Information Feeding and Privacy Protection: Culture-based Considerations for Filtering and Safeguarding Information in Online Social Networks

Arunee RATIKAN

Japan Advanced Institute of Science and Technology

Information Feeding and Privacy Protection: Culture-based Considerations for Filtering and Safeguarding Information in Online Social Networks

by

Arunee RATIKAN

Submitted to
Japan Advanced Institute of Science and Technology
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Supervisor: Professor Mikifumi SHIKIDA

School of Information Science
Japan Advanced Institute of Science and Technology

December, 2014

Abstract

Online Social Networks (OSNs), i.e. Facebook, Google+, Twitter, etc., currently play an important role in communication and social interaction. The main goal of OSNs is to provide an available space where the user, or creator, can create and post information. At the same time, this information is consumed by other users, or readers, anywhere at any time. OSNs bring the users closer even though they may be of different race, religion, or gender and living in different places around the world. Therefore, they are viewed as tools for communication, and social interaction. However, when the number of users with different cultures in OSNs increases, the variety, amount, and sensitivity level of information also increases dramatically. Accordingly, users in OSNs are facing many problems caused by cultural differences, such as information overload, loss of privacy, misunderstanding information, in-group-out-group bias, and so on.

This research focuses on reducing the problems of information overload, over consumption and loss of privacy in information sharing by consideration of the cultures in OSNs. This is because information is an important mediator in communication and social interaction among users in OSNs. Furthermore, it is presently sent and received by users without consideration of cultural differences. Thus, the users, both readers and creators, cannot fully take advantage of OSNs. For example, the users receive too much information, and lack the ability to control that information.

For the consumer, the information overload problem originates from different cultural patterns of text, large numbers of users in OSNs, and the large amount of information. This problem causes the readers difficulty in finding interesting and important information, creating feelings of confusion, anxiety, and annoyance when readers consume excessive information on the Social Network Page (SNP). Consequently, readers cannot adequately consume high-quality information. To address the information overload problem, cultural differences in information consumption are investigated by using a survey.

A set of influential features and factors is additionally prepared to filter the information based on cultural differences. Subsequently, a new type of Information Feed Mechanism (IFM) is proposed. It considers the cultural differences in selecting interesting information to appear on the reader's SNP. The proposed IFM helps the readers to consume interesting information within a short period of time. The most suitable information, based on the reader's current situation and nationality is served to them, which is good for both businesses and readers. The proposed IFM can be applied to very large data sets, providing exponential growth in OSNs by using a parallel concept. The analyzed results of the proposed IFM can also be adopted by many countries, societies, and businesses.

For information sharing, the loss of privacy is a crucial problem in OSNs due to lack of collective privacy protection. Only the owner who creates and posts the collaborative information on OSNs, can control it. The co-owners who are associated with the collaborative information, might lose privacy from tagging, mentioning, or sharing such information posted by the owner without asking their permission. Moreover, the co-owners might not realize their information is being managed by others. It is possible that collaborative information might leak to unwanted target readers. To balance the need for information sharing and the privacy protection of the owner and co-owners, a Collective Privacy Protection (CPP) is proposed. The concept of majority vote is applied to the proposed CPP. The co-owners can make a decision whether or not to allow this collaborative information to be posted on OSNs by consideration of the privacy policy. The proposed CPP identifies the privacy conflicts between the owner and co-owners and provides a suitable solution for those conflicts, although only one co-owner can reject the privacy policy. By using the proposed CPP, the owner and co-owners share the collaborative information on OSNs with less privacy concerns, because the collaborative information will not leak to unwanted target readers. The proposed CPP encourages the owner to take responsibility for the co-owners' privacy by asking permission. The proposed CPP brings about negotiation of privacy, based on the cultures of the owner and co-owners by asking their permission. This indicates that many factors influence the co-owner's response, such as power distance, individualism vs. collectivism and so on. In addition, the proposed CPP helps reduce the crime problems in society, e.g. robbery, defamation, and violation of

portrait rights.

Based on this research, the variety, sensitivity and large amount of information coming from users of different cultures cause information overload and loss of privacy will be controlled, reducing stress for the users. The users can take full advantage of information consumption and the information sharing in OSNs.

Keywords: Cultural differences, Information feeding, Privacy protection, and Online Social Networks

Acknowledgments

First of all, I wish to express my sincere gratitude to my principal advisor, Professor Mikifumi Shikida of the Japan Advanced Institute of Science and Technology (JAIST). My advisor encouraged, gave valuable advice, and provided unlimited support to me during my research at JAIST. He devoted time to teach me not only in relation to my Ph.D research but also about life after graduation. Moreover, he always shows concern for my health and feelings, and continually motivates me. I am so grateful for the excellent supervision of Professor Mikifumi Shikida, without whose help, I would not have been able to complete this research.

I would like to thank Professor Yoichi Shinoda, Professor Susumu Kunifuji, Associate Professor Takaya Yuizono, and Professor Takashi Yoshino for the valuable guidance offered by them in my research. Their guidance has helped to improve my research considerably. I would like to express my thanks to my minor research supervisor, Associate Professor Kiyoaki Shirai, who provided me with the knowledge of natural language processing. The time I spent with him during my Ph.D research was invaluable and he often provided small lectures to guide me.

I would like to show my gratitude to my professors in Thailand, Professor Dr. Thanaruk Theeramunkong, who recommended me to study at JAIST and always supported me at all times. I gratefully acknowledge Associate Professor Dr. Waree Kongprawechnon, Assistant Professor Cholwich Nattee, and Dr. Denduang Pradubsuwan. All of them gave me a lot of support and encouraged me in my Doctoral degree in Japan. They were always on hand to help me.

I offer my sincere thanks and appreciation to all of my friends from Thammasart University, Chulalongkorn University, and JAIST. Dr. Kobkrit Viriyayudhakorn, Ms. Chompoonoot Veerakiatikit, Ms. Pimnapa Atsawintarangkun, Ms. Pimolluck Jirakunkanok, Mr. Chaianun Damrongrat, Ms. Lin Xiaohui, Ms. Ikue Osawa, Mr. Sai Ryo, and other Thai, Japanese, and Chinese friends provided help and suggestions. Although they were busy, they still found time for me, and we always helped and advised each other. I am

very fortunate to share my life in Japan with the many friends I have made from countries all over the world. I would like to thank the members of the Shikida laboratory: Mr. Yukinori Sakashita, Mr. Ren Sasaki, Ms. Kaori Ishii, Mr. Toshiki Kawai, and Mr. Hiroyuki Nishino who have looked after me during my stay in Japan. Whenever I have a problem with the Japanese Language, they help me. I am also indebted to my Japanese teachers, Kuniko Kitajima sensei, Hisami Matsuda sesei, Yuko Hosotsubo sensei, and Sukeda sensei. Since my first Japanese class, I have very much appreciated their time contribution and valuable advice, particularly regarding Japanese culture, and language.

I would like to sincerely thank my lovely family for their encouragement and huge support. When I am sad, they comfort me and help me to relax and not take myself too seriously. Although they were unable to offer any practical help in my Ph.D research, their enormous contribution to my personal well-being has been paramount. I must also give special thanks to Dr. Konlakorn Wongpatikaseree who devoted his time to help, care and support me, whatever the circumstances, and I know we will always continue to encourage each other.

Finally, I would like to express thanks to JAIST for giving me the opportunity to study under the Doctoral under Graduate Research Program (GRP) by offering financial support, and the staff were always kind and helpful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

"Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god."

–Aristotle

Social life is necessary for humans. In other words, humans cannot live without society. When a group of people live together, communication is a mediator for transferring information from one person to another.

Many years ago, there was no communication technology, therefore conveying information from one to another was difficult. As shown in Figure 1.1, people used drums and smoke signals as a way of communicating with others. Later, the means of communication was adapted. A written message was delivered to a recipient by using carrier pigeons or pony express. However, communication over long distances with these methods might not be efficient. This is because it took too much time and the letter might perhaps be lost or damaged during the delivery process. Thus, communication occurred between particular groups of people living in the same or nearby. In other words, the cultural exchange via

communication in the past was much more difficult than it is in the present.

Communication technology was improved with the introduction of the telephone, making it more convenient than in the past, enabling people to communicate with each other over longer distances, providing them with the opportunity to contact others living in different places or countries. When distance is no longer a barrier to communication, cooperation among countries can be established, for business, military, medical, and so on.

Since the Internet was first introduced in the 1960s, it has become an important role in daily life. At present, we have the convenience of communications such as electronic mail (e-mail), instant messaging, the World Wide Web and so on. The Internet has many advantages for communication in that it reduces costs, saves time, and improves life quality. For example, we do not need to have a meeting in the same room, but we can discuss via the Internet. This indicates that no matter where we are in the world, we can communicate easily.

In recent years, Online Social Networks (OSNs), such as Facebook [10], Google+ [11], and Twitter [12] have become significant communication media. They change the communication style so that conveying information from one to another does not need to be by direct communication such as e-mail or telephone. Moreover, users of different cultures around the world can easily communicate with each other. This is because OSNs

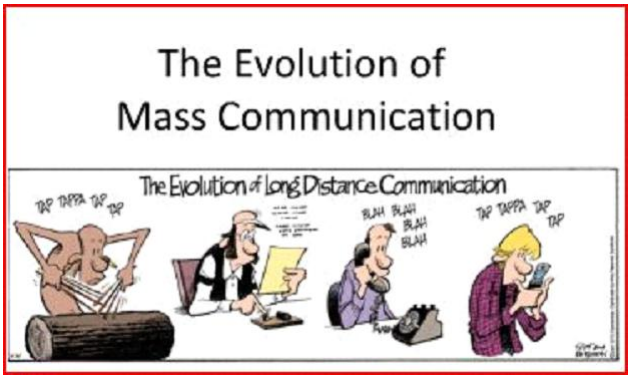


Figure 1.1: Evolution of communication [1]

[13] bring people of different races, age groups, and cultures closer together by enabling OSN users to consume unlimited information (i.e. news updates, friends' stories, events, etc.). In addition, users can share any information (i.e. text, photo, video, link, etc.) with others in real-time anywhere at any time. The users meet each other in OSNs for a variety of purposes (i.e. entertainment, business, relationship maintenance, etc.).

Furthermore, OSNs have a social impact on various aspects, such as marketing, business, charity donations, politics, and so on. For instance, when taking part in a social game on Facebook, the player has an opportunity to earn points from the game and turns these points into a charitable donation. This donation can provide drinking water and meals for children in Haiti [14]. Moreover, OSN might change user behavior. For example, some users spend more time reading the news via OSNs instead of in a newspaper. Without doubt, OSNs have rapidly grown with a large number of people in many countries, such as America, Russia, Japan, and so on. The number of users in OSNs has increased dramatically, and there are more than 100 million active users [15]. Especially Facebook, where the monthly active users numbered 1.19 billion in September 2013 [16], and is now the most popular communication media as illustrated in Figure 1.2.

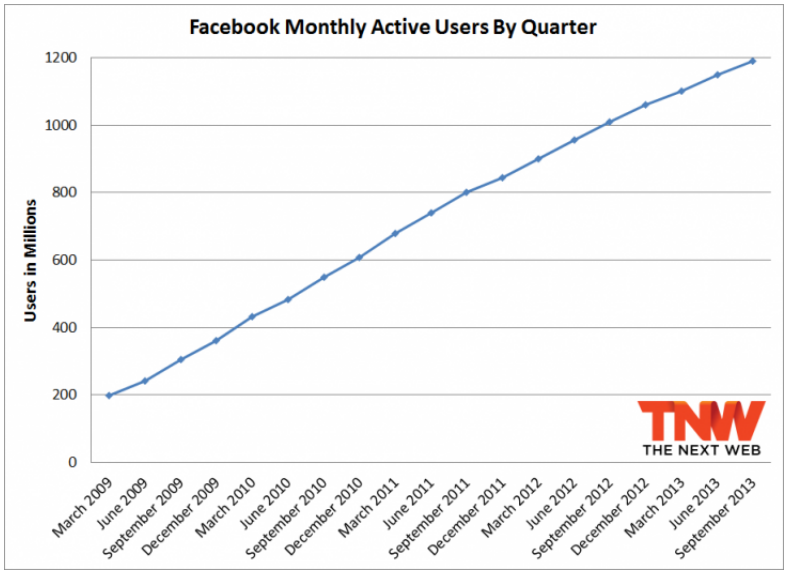


Figure 1.2: Facebook monthly active users [2]

Generally, users in OSNs can carry out two main communication processes: informa-

tion consumption and information sharing. Information consumption refers to a process that the user, or reader, has the ability to consume any kind of information via a Social Network Page (SNP), such as on a wall in Facebook, Home Stream in Google+, Timeline in Twitter, etc. Information sharing is a process that the user, or creator, creates and posts information on OSNs. The information can be via text, photo, video, or link. When the information is posted on OSNs, it is quickly distributed and monitored by other readers.

In this research, the focus is on reducing problems by the consideration of cultures in OSNs. Nowadays, OSNs have a major impact on communication and social interaction, and are widely used in many countries. When there is a large number of users and amount of information increases in OSNs, the users face many problems, such as information overload, loss of privacy, and so on. These problems may relate to cultural differences because users in each culture do not live together in the same place, share the same experiences, and might have different preferences. Therefore, reducing these problems by considering cultural differences in OSNs is not an easy task.

1.2 Problems related to the cultural differences in OSNs

The users in OSNs come from many countries around the world as indicated in Figure 1.3. They can easily communicate with each other via OSNs. When users from different cultures or countries take advantage of OSNs for communication and social interaction by consuming and sharing information, the variety, quantity, and sensitivity levels of information increase dramatically. This information can lead to many problems for the users, such as information overload, loss of privacy, misunderstanding of information, in-group-out-group bias, and so on. Each problem causes difficulties for users of different cultures in OSNs because each user grows up in a particular culture, learning the language, experience, and rules from that culture. A summary of the problems are described as follows:

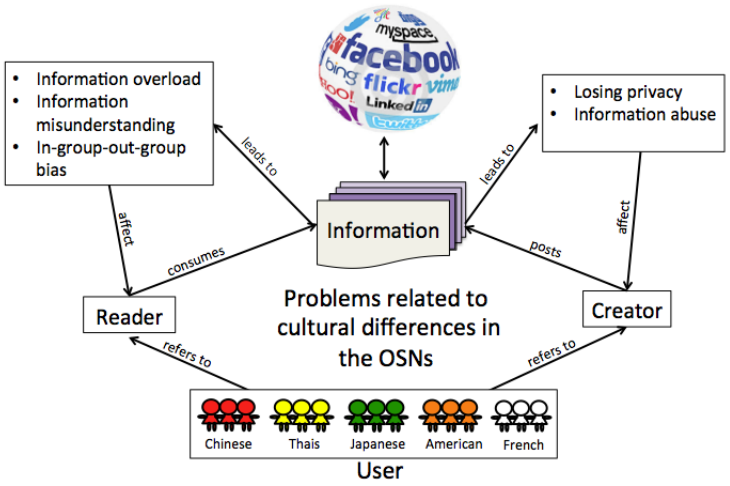


Figure 1.3: The problems related to the cultural differences in OSNs (taken from www.dreamstime.com)

1. **Information overload**

   There is excessive amount of information contained on the reader's SNP. The reader receives too much information and does not want to know all the information at the same time. He/she has difficulty with finding interesting information to focus on because it is blocked by other information. Moreover, this information, which is shown on the SNP, does not relate to the reader's culture. As a side effect, the reader feels confusion, anxiety, and dissatisfaction [17][18]. This emotion can lead to activity reduction in OSNs [19], such as friend blocking, relationship break-up, and limitation of friend requests. The information overload problem derives from three causes: cultural pattern of text, the number of users in OSNs, and the amount of information in OSNs.

   In the cultural patterns of text, the reader might be unfamiliar with the writing style, language, and complexity of content [20]. Each culture has a different pattern of text. When readers of different cultures are the creators, they have to cope with this unfamiliarity. Therefore, the readers can easily experience information overload.

   As for the number of users in OSNs, nowadays these increase continually year by

year. For instance, Facebook had 1.19 billion active users in September 2013 [16]. The users come from many countries around the world and have an opportunity to create and post an enormous quantity of information on OSNs.

To increase the amount of information in OSNs, the creators can portray themselves as publishers or generators of the information. Therefore, they can create and post any kind of information anywhere at any time. This can increase the amount of information. For example, Facebook served 2.5 billion photos per week in September 2013 [21].

2. **Misunderstanding of the information**

   This problem occurs when readers misunderstand the intention of the creator. It can happen in all areas of communication even when the creator and reader come from the same culture. The readers and the creators with different cultural backgrounds are more likely to misunderstand than those with the same cultural background. In OSNs, the creators and readers generally have a diversity of cultures, and do not communicate face-to-face. This indicates that readers can easily experience this problem because they do not have the same background or life style and way of thinking as the creator. Sometimes, the text is not clear enough. For example, the creator may intend to post a story as a joke but the reader with a different culture might not clearly understand it. Hence, that story may not be funny to them.

3. **In-group-out-group bias**

   In social groups, people tend to give preferential treatment to members of their own group. This treatment differs from the members, who are viewed as being outside the group. This bias depends on culture, language, gender, and so on [22]. In OSNs, the reader's SNP can contain a mix of information, which is posted by many creators. It is possible that the reader selects only sections of information from some groups of creators, whom the reader accepts as members in the same group. Thus, the information posted by other groups of creators might be ignored although it is interesting or important. For instance, a Japanese creator may post texts about earthquakes and safety tips but a Thai reader might skip this information because

of the language used, affinity, and so on.

4. **Loss of privacy**

   In OSNs, when the creator posts information, this information is frequently viewed as being owned by this creator, or owner. In the real world, collaborative information (e.g. texts, photos, video, links) is associated with multiple users, known as co-owners. For example, Alice, Bob, and Carol took a photo together. The photo is considered as collaborative information. Even though OSNs allow the creator to post any information, it has not provided an adequate mechanism of privacy protection for the creator [13][23]. Thus, the creator (the owner and co-owners) might lose their privacy. The loss of privacy can be caused in three possible ways.

   Firstly, the co-owners do not have permission to control the collaborative information with which they are associated. In addition, they might not realize that their information is being managed by others. The owner can normally tag, mention, or share the collaborative information with the co-owners without asking permission. For example, the photo containing the owner and the co-owners is posted by the owner without permission from the co-owners. Although the owner has the right to control publication of this photo, the co-owners should be asked for permission because they also have portrait and photo rights. Therefore, the information might leak to unwanted target readers (with whom the owner or co-owners do not wish to share).

   Secondly, sensitive information levels rely on cultural differences. In other words, sight of the same information can lead to different privacy concerns. For example, the owner from a Western country may view a photo of people kissing in public as nothing special, but the co-owners from Asian countries could think that it is shameful and should not be spread on OSNs. Moreover, each culture has a different style in information sharing. For instance, Hong Kong creators are more likely to disclose personal information, while French creators feel less in control when updating personal information [24].

Lastly, negotiation of privacy based on culture is a difficult task because the real meaning of the owner and the co-owners of different cultures needs to be understood. This negotiation is not done via face-to-face communication, and hence when the co-owners are in different places to the owner, the owner has no clue as to the real meaning which the co-owners are trying to express.

1.3 Statement of problems

This research aims to reduce *information overload and loss of privacy problems* by consideration of cultures in OSNs. These problems are serious because the information is sent and received by users in OSNs without consideration of cultural differences, despite users in each culture having their own way of thinking, feelings, and behavior based on their experiences. Accordingly, both readers and creators cannot take full advantage of OSNs. For example, the users receive too much information and lack the ability to control that information.

In information consumption, the reader is provided with all kinds of information by the Information Feeding Mechanism (IFM) [25][26] via SNP. Nonetheless, the reader cannot adequately consume high-quality information because the there is too much information on the SNP, and it is not consistent with the reader's culture. The information overload problem has been addressed by various approaches. Lops et al. [27] and Vanetti et al. [28] have tried to filter uninteresting and unimportant information by using the reader's interests. However, this might not be enough. Paek et al. [29] use almost all available data in OSNs as the features. However, these features cannot achieve high classification accuracy. This indicates that using a large number of features might not be the optimal solution for filtering uninteresting and unimportant information because some features do not influence the reader. From previous works, selected features were used for information filtering without considering the cultural differences of the reader. Each feature used to filter the information in each domain might be different. Therefore, it is not an easy task to find the necessary features appropriate for each individual reader's culture.

For information sharing, loss of privacy is a crucial problem in OSNs. Generally, the

collaborative information created by the owner and co-owners can be controlled only by the owner. When the collaborative information is posted on OSNs without asking the permission of the co-owners, they might not realize that their information is being managed by others. In addition, they might not want to post this collaborative information because at least one co-owner wishes to keep it private. If the collaborative information leaks to unwanted target readers, it can lead to loss of privacy and other consequential problems, such as theft, defamation, infringement of portrait or photo rights, and so on. This indicates that inefficient privacy settings are provided by OSNs and different privacy preferences depend on culture. Therefore, it is hard to control collaborative information. Many approaches [30][31] try to protect privacy, but they still allow only the owner to control privacy management. In some research studies [13], the owner and co-owners can create the privacy policy, but it can lead to management problems because it cannot satisfy everyone. When every co-owner agrees to a privacy policy, the information lacks freshness and interest. This process is very time consuming.

1.4 Research objectives

To address these shortcomings, the main propose of this research is to build a knowledge base for information feeding and privacy protection with consideration of cultures in OSNs. This knowledge base is consistent with the user's requirements in consuming and sharing information. Therefore, to this end, four sub-objectives are addressed as follows:

1. This research should provide an analysis of cultural differences in order to provide basic knowledge for creating a new type of IFM in OSNs. The term cultural differences in this research refers to the study of thinking patterns, feelings, and beliefs that come from the traditions of each country.

2. OSNs should possess a new type of IFM for solving information overload taking into consideration cultural differences. The proposed IFM must have a module to filter uninteresting information not consistent with the reader's current situation and preferences.

3. Balancing the need for information sharing and privacy protection is necessary and

important in OSNs; not only for the owner of the information but the co-owners also need to have their privacy protected.

4. This research tries to encourage the owner to take responsibility for the co-owners' privacy although the co-owners may not have the same culture as the owner. In addition, the co-owners should have the right to decide which information should be posted on OSNs in order to avoid loss of privacy.

1.5 Research methodologies and originalities

In order to achieve the above objectives, this research is composed of three main processes which impact on cultural differences in information feeding, culture-based preference for the IFM, and Collective Privacy Protection (CPP) using majority vote. The first two methodologies are for alleviating the information overload problem, while the last one is for resolving the issue of loss of privacy.

1. **An impact of cultural differences for information feeding**

   This is a process that studies and compares the culture of each country by using features and factors for evaluation. In this research, the terms of feature and factor generally mean that an attribute influences the reader's behavior, thoughts, and feelings in selecting the information to read. More details of these terms will be described later. For this process, it consists of two main components: data collection and data analysis.

   The first component, *data collection*, involves gathering data from a survey. It requires two sub-components: feature construction and data pre-processing. Feature construction is the definition and utilization of many types of features for information filtering. However, these are still investigated in order to find influential features of the readers, such as type or group of relationship between the reader and the creator, information category, the reader's current situation, and so on. Data pre-processing is carried out because after obtaining results of the survey from the respondents, there is a need to make sure that the collected data is ready to use. If

the data is of poor quality and used for other components, satisfactory results will not be obtained.

The second component; *data analysis*, selects the features which influence the readers in order to use them in information filtering, as is described in Chapter 4. Nonetheless, based on the survey results, the contributory factors, reflected in the answers, also impact on the readers, such as age group, career, gender, and so on. Hence, this component compares cultural similarities and differences between Japanese and Thais by using two sets of features and factors.

The originality of this research involves studying cultural differences in consuming the information domain of OSNs, which has not been emphasized by previous research works. Reducing information by considering the reader's culture can alleviate the information overload problem [20] since the selection of the reader will be relevant to their requirements.

2. **Culture-based preference for the IFM**

This process is aimed at developing a new type of IFM to attempt to solve information overload using the consideration of culture dependency. The concept is that information should be dynamically selected and displayed based on the reader's current situation in order of the reader's preference. To do this end, this research uses the survey results from the first process as a training set. To filter less interesting information, this research uses the concept of classification. Moreover, three classification algorithms are investigated: Decision Tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB), using classification accuracy and time complexity as measurements. Furthermore, the performance of the information filtering is compared with virtual existing information filtering. After that, an experiment is conducted using real data and evaluating the proposed IFM using Japanese and Thai respondents.

This research is novel because the proposed IFM selects and displays the information

according to the reader's culture. Although several research works have attempted to solve the information overload problem by serving relevant information to the reader in OSNs, they have used some features for filtering without consideration of culture, which may not be efficient. The classification algorithm used in the proposed IFM is a fast and highly scalable building model. The performance of the proposed IFM is reliable because it uses real data and is tested by Japanese and Thai respondents. In addition, the proposed IFM can be applied to other cultures, societies, and businesses.

3. **Collective privacy protection by using majority vote**

The CPP is proposed to balance the need for information sharing and privacy protection for the owner and co-owners. The concept of this process is that it enables the owner to create a privacy policy for sharing collaborative information and co-owners to participate in the privacy policy by vote (acceptance and rejection) whether or not this collaborative information should be posted on OSNs. This indicates the right of co-owners to collaborative information. If the vote results in acceptance of more than half, this means the collaborative information can be posted to OSNs. The proposed CPP additionally identifies privacy conflicts and provides a suitable solution for those conflicts. It still supports the co-owners who reject the privacy policy because privacy is considered as high priority. This research analyzes the factors which help protect privacy and investigates the opinion of co-ownership. Furthermore, the performance of the proposed CPP is compared with other techniques.

The originality of this research is that it attempts to protect the privacy of not only the owner but also co-owners associated with the collaborative information. The co-owners should have the right to decide on their information. The co-owners from each culture have different privacy preferences because they have different sensitivity levels. However, a few research works [13][23] recognize the issue of co-owners' privacy. The proposed CPP helps the co-owners to acknowledge that their information is being managed by others because OSNs allow the owner to tag, mention, and share collaborative information without asking permission from the

co-owners.

1.6 Chapter organization

This research is divided into eight chapters. Chapter 1 introduces this research with a statement of the issues, research objectives, methodologies, and originalities. Chapter 2 describes backgrounds to various research and related works. It starts with an explanation of OSNs, including cultural differences in the IFM and privacy protection models. Chapter 3 provides the architecture of this research and its contribution. Chapter 4 focuses on the investigation of cultural differences in information feeding. This chapter explains data collection and analysis. Thereafter, a new type of IFM is proposed and presented for culture-based preferences in the information feeding mechanism in Chapter 5. In Chapter 6, the CPP is proposed to balance information sharing and privacy protection for the owner and co-owners. Chapter 7 shows contributions to this research in terms of social and academic impact. Finally, Chapter 8 describes the conclusion of this research and gives future direction.

Chapter 2

Background and literature review

This chapter describes the characteristics of Online Social Networks (OSNs), their cultural differences, Information Feeding Mechanisms (IFMs) and privacy protection models. OSNs are introduced by indicating famous examples like Facebook, Google+, LinkedIn, and Twitter, and their important points. The cultural differences are explained by the definition, nationality, and related works. The IFMs in OSNs are then described using existing IFMs from Facebook, Google+, and Twitter. Causes of the information overload problem and literature review for solving it are then presented. Finally, privacy protection models in OSNs are provided to explain why the users feel a loss of privacy when sharing information and how the privacy problem is solved by previous research works.

2.1 Overview of OSNs

When Web 2.0 [32] was initially launched, it was considered to be a website for the next generation, providing services to allow users to connect each other. OSNs (referring to online communities whose main goal is to make available an information space, where each social network participant can publish and share information defined by [33]), blogs and wikis are now popular with several groups of people, not just the young, but also adults and the elderly. Companies, organizations, job seekers, and other individuals are able to obtain useful information from their participation on OSNs, such as Facebook, Google+, LinkedIn, Twitter and so on. This is because OSNs like User-Generated Media (UGM) [34] enable the user to share information so that the other users can take advantage

by consuming it. For example, the user can share personal stories, interests, activities, services, and so on. Therefore, OSNs are considered to be a new tool for social interaction and communication [35]. Recently, OSNs have predicted that they will continuously grow with more diverse populations. In other words, many users from different age groups, careers, or places in the world will use OSNs for many purposes. This brings the users closer, gaining a great deal of knowledge from others' experiences. However, OSNs have an effect on sociability in negative ways. For instance, the users spend much more time interacting with each other via the Internet than face-to-face [36].

2.1.1 Examples of OSNs

This sub-section describes examples of well-known OSNs. Facebook, Google+, LinkedIn, Twitter are selected for examples due to their volume of active users [37]. The active user is an important metric as it refers to the ability to connect with others in OSNs, but it cannot represent their size.

1. **Facebook** was introduced in February 2004 by Mark Zuckerberg and his team [32] [34]. The Facebook users can share any information (e.g. text, photo, video, or link) with other users via web browser or mobile application as shown in Figure 2.1, numbers 1 and 3. Facebook allows the user to interact with friends via a comment or pressing the like button on a friend's information page, as well as the chat feature through News Feed. "News Feed" is a page where it shows highlighted information about other users with whom that user has a relationship, as illustrated in Figure 2.1, number 2. The information appearing on the News Feed can be status updates, events, or birthdays of other users. Usually, the Facebook user can create profile (i.e. name, e-mail, phone number, photo, education, etc.), list of interests (i.e. music, movies, sports, etc.), friend list (referring to friend request, making a group). In the third quarter of 2013, Monthly Active Users (MAUs) totaled 1.19 billion as of September 30, 2013 [16].

2. **Google+** managed by Google was launched in 2011 and is currently the second-largest OSN in the world. Google+ allows the user to selectively share and consume information via a specific "Circle" in his/her Home Stream as presented in

Figure 2.2, numbers 1 and 2 [35][38]. Home Stream shows the information either specific or publicly shared, according to time. However, the user can filter the information by selecting specific circles. Until recently, the user could control the amount of information appearing on the Home Stream by adjusting the quantity (more, standard, and fewer) in each circle, including what's hot and communities. This indicates that Google+ no longer shows all information in the user's Home Stream. In addition, Google+ provides the Hangout feature, which is a group video. This feature can support a maximum of 10 users for one hangout. In 2013, Google+ had 540 million active users [37].

Figure 2.1: Facebook News Feed

3. **LinkedIn** was started in 2003 and is the largest and most popular OSN for people in professional occupations [32]. LinkedIn allows the users to make a contact list and profile (using personal and professional data) as shown in Figure 2.3 and maintains relationships with those in their contact list [27]. This is known as a connection. One user can also invite another to become a connection. However, the invitee can reply through the system to indicate that he/she does not know the inviter or the inviter is recognized as spam. One of the main features of LinkedIn is that it can be used for finding jobs. In other words, a user on LinkedIn can be recommended to an employer, and the employer is able to check the profile of potential candidates.

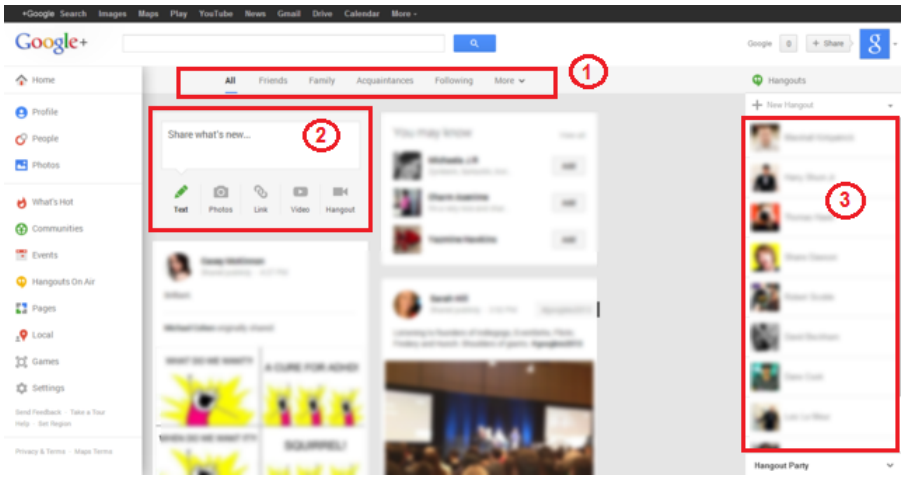


Figure 2.2: Google Home Stream [3]

Nevertheless, the user can press the like button or congratulate other users in their new job. Presently, LinkedIn provides 20 languages. In 2013, it had 259 million active users [37].

4. **Twitter** was launched in 2006. Twitter [39][40] uses a concept of social networking and SMS messaging enabling the user to publish a "tweet" (message) with a maximum length of 140 characters. It provides a simple way to indicate another's status via Twitter Timeline, where it shows tweets in reverse chronological order as demonstrated in Figure 2.4. This means that a new tweet is added at the front. The user can see the mix of tweets in the Timeline by following the people involved. Twitter not only allows the user to see all of the tweets on his/her Timeline but it also provides a search engine to filter out unwanted or unrelated tweets and discover new ones. One user can follow any other but the user who is being followed does not need to follow back. A follower on Twitter indicates that the user receives all tweets from those who the user follows. Within the tweet, Twitter provides the user with a well-defined mark-up culture: RT stands for "retweet", a command to allow the user to spread information, "#" known as a "hashtag" is followed by a keyword like a tweet category and "@" followed by a user identifier address. In 2013, Twitter had 232 million active users [37].

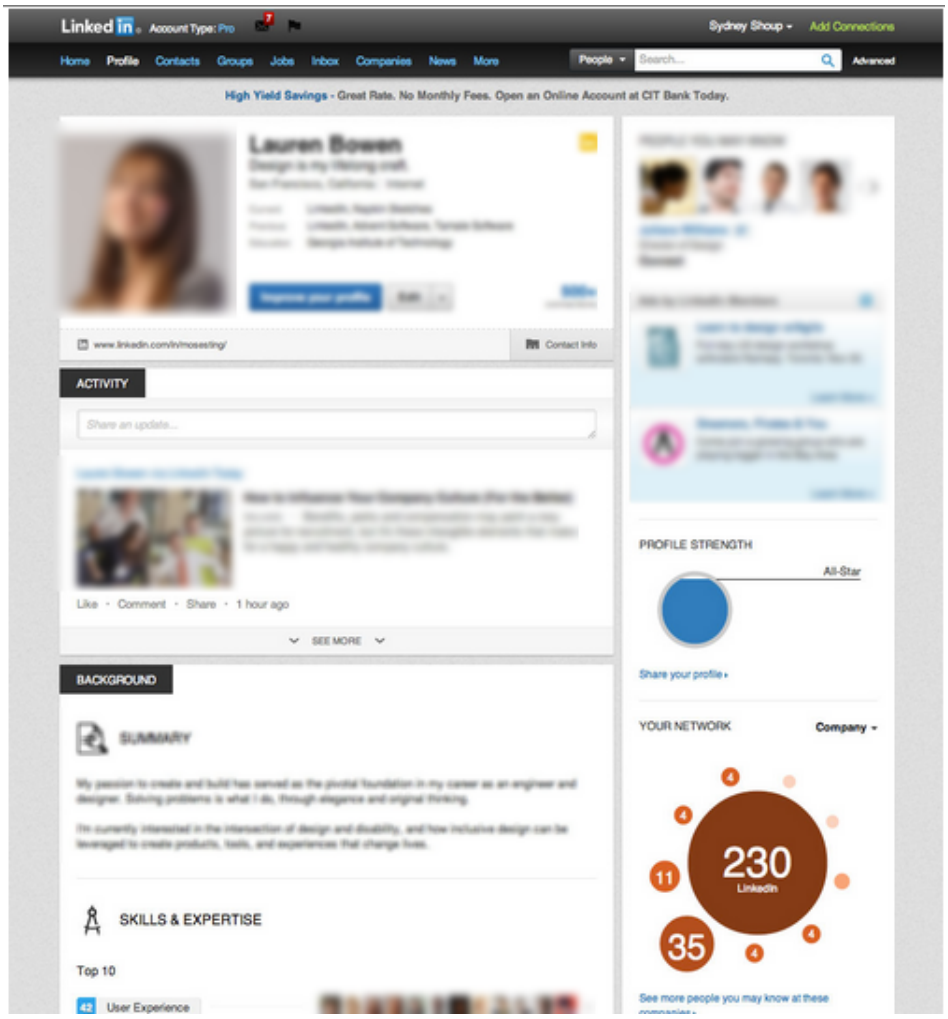


Figure 2.3: LinkedIn Home page [4]

2.1.2 The importance of OSNs

1. OSNs give users an opportunity to freely share their feelings, opinions, or daily life in several areas; such as politics, fashion, economics, and so on. The users also receive feedback from other users via comment, pressing the like button (in Facebook), or +1 button (in Google+). This indicates the user is available for social interaction and communication with others anywhere at any time via OSNs.

2. OSNs bring the users closer via social interaction and communication although each person may be at a different place in the world. The users can learn about each other, including their culture, values, customs, and traditions with location posing no barrier.

Figure 2.4: Twitter Timeline

3. OSNs are viewed as tools for relationship maintenance. Many users occasionally meet old friends who have moved away, and the user may not have seen them for a long time but want to continue to maintain the relationship.

4. Because OSNs can propagate any information from long distances, very quickly, they impact on society in several ways, i.e. social welfare, medicine, politics, etc. For instance, OSNs can be used for business. If companies want to advertise their products to thousands of people, OSNs provide a good strategic vehicle to enable the companies to achieve their goal without cost.

2.2 Cultural differences in OSNs

Hofstede [41] explains culture as always being a collective phenomenon. This is because people who live together in the same social environment will develop cultural knowledge.

This indicates that cultural knowledge comes from one's social environment, it is not genetic. Therefore, understanding the thoughts, feelings, and actions of people in every culture around the world is challenging and interesting.

2.2.1 Definition of culture

The word "culture" has many meanings, but its origin comes from the Latin [42] meaning: tilling of the soil [41]. In the Western language, "culture" is given the meaning of civilization or refinement of the mind. The definition of culture has been widely discussed in research, but there is confusion amongst anthropologists and sociologists as to the definition. This causes the definition to ranges from very complex to very simple, thereby clarifying the content and boundaries of culture as necessary.

- "Culture consists in patterned ways of thinking, feeling, and reacting acquired and transmitted mainly by symbols, constituting the distinctive achievements of human groups, including their embodiments in artifacts; the essential core of culture consists of traditional (i.e. historically derived and selected) ideas and especially their attached values." was presented by Kluckhon in 1951 [43]. This definition is often cited in many research works.

- "Transmitted and created content and patterns of values, ideas, and other symbolic-meaningful systems as factors in the shaping of human behavior and the artifacts produced through behavior" was proposed by Kroeber and Parsons in 1958 [44].

- "By culture we mean an extrasomatic, temporal continuum of things and events dependent upon symboling" was given by White in 1959 [45].

- Culture is shared mental software, "the collective programming of the mind that distinguishes the members of one group or category of people from another". This definition was introduced by Hofstede in 2001 [46]. He also explained further that the group or category in the definition could not only relate to a national society but also regions, ethnicities, age groups, genders, etc.

Based on the four definitions above, the meaning depends on the context of where the definition is applied or used and hence it is difficult to create one single definition. This

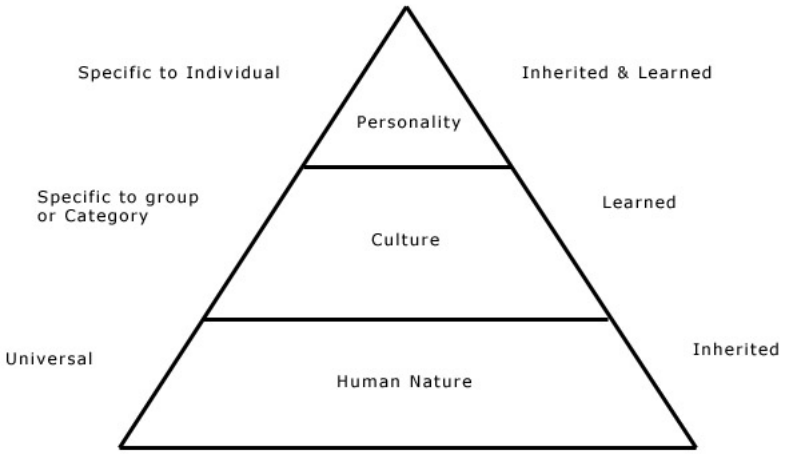


Figure 2.5: Three levels of uniqueness in mental programming

research follows the cultural definition introduced by Hofstede. From his definition, he attempts to compare the patterns of individuals by using the analogy of the way a computer is programmed. Every person has his or her thinking pattern, feelings, and actions known as *mental programs*. The source of a person's mental program is taken from their social environment. This source starts with the family, school, workplace, community, and so on. The mental program of each person relies on the social environment in a different way. For instance, a person who has lived or worked in a slum will have a different mental program from those who have not lived there. Hofstede also expresses that culture should be considered separately from human nature and personality as illustrated in Figure 2.5.

Human nature refers to the characteristics that all humans usually have. It indicates the universal way which humans are able to think, feel, and act, together with their need to interact with others independently and naturally. Human nature is inherited genetically, like the operating system of a computer, thereby making the human being what he/she is.

Personality means that humans have unique characteristics of mental programs that they share with others. Some part of the personality comes from a unique set of genes, and learned from unique personal experience.

Moreover, Hofstede attempted to distinguish national cultures from each other by analyzing the employee value scores gathered within IBM. It was statistically divided into four dimensions, known as the Hofstede dimensions of national culture. The four dimensions are Power Distance (PDI), Individualism versus Collectivism (IDV), Masculinity versus Femininity (MAS), and Uncertainty Avoidance (UAI). Subsequently, in 1991, Michael Harris Bond added the fifth dimension, Long-Term Orientation (LTO). However, the collected data is replicated and extended for analysis until 2010, and the data collection came from 76 countries [47]. Recently, the sixth dimension; Indulgence versus Restraint (IVR), was added into Hofstede dimensions by Michael Minkov. A summary of Hofstede dimensions is shown in Table 2.1.

Table 2.1: Hofstede dimensions

Hofstede dimension	**Definition**
PDI	The degree to which the members of a society are accepted unequally.
IDV	The degree to which a society focuses on individual or collective achievement and interpersonal relationships.
MAS	The degree to which a society is depended on achievement, assertiveness and competitiveness.
UAI	The extent to which the members of a society feel uncomfortable, with uncertainty and ambiguity.
LTO	The extent to which a society shows a pragmatic future-oriented perspective rather than short-term point of view.
IVR	The degree to which a society allows relatively free gratification of basic and natural human drives related to enjoying life and having fun.

2.2.2 National culture and Asian culture history

Studying the cultural differences by using nations as units of analysis is not acceptable [48][49][50]. Some of them [51] state that national borders are not enough for analysis because each nation has sub-cultures. Moreover, the concept of culture is applied to society, not to the country [41]. Society refers to an organization or group of people gathering for a particular purpose or activity. Nonetheless, each nation makes own form, which is de-

veloped for people in the nation, although there are different sub-cultures. This can create unique national cultures. Nowadays, national statistical data, e.g. GDP per person, and economic growth rates is collected to indicate values, norms, and beliefs. Furthermore, using the nation as a criterion for study the cultural differences can lead to cooperation in aspects such as business, education, and politics.

Our world consists of seven continents; Asia, Africa, North America, South America, Antarctica, Europe, and Australasia. People in the same continent are culturally different from those in other continents. For instance, the communication style of North Americans completely differs from that of Asians. People on the same continent share similarities due to a similar landscape or their proximity to each other. However, people of different nations on the same continent still have different communication styles, language, beliefs, etc. because each nationality will adapt themselves to the environment, where they have grown up and lived, via social learning. The same rule or law can be applied to specific nations but they cannot adapt to others because of cultural differences.

Asia is a good example of cultural variety, such as the Mesopotamian civilization in East Asia, Chinese civilization in East Asia, the Indian civilization in South Asia, and so on. Although these civilizations have spread and been exchanged with other nations and regions in Asia, e.g. mathematics, law, innovation of technology, and belief, each nation still preserves its own culture and mixes these civilizations with its own culture. However, language was developed individually by each nation. Confucianism is an ethical and philosophical system. It was developed by the Chinese philosopher Confucius and influences the lifestyle and art of Asian people like Chinese, Japanese, Korean, and Singaporean. The teaching of Confucius is about practical ethics, which describe filial piety, respect for elders, social obligations, and rules of courtesy for daily life [52]. There are four key points in the teachings of Confucius [41].

- The stability of society is based on unequal status relationships between people.
- The family is the prototype of all social organizations.
- Virtuous behavior toward others consists of treating others as one would wish to be treated oneself.

- Virtue with regard to one's tasks in life consists of trying to acquire skills and education, working hard, not spending more than necessary, being patient, and persevering.

Moreover, there are many interesting cultures in Asia. As with the Middle Asia, South Asia has not been directly influenced by the teachings of Confucius. For example, India comprises several regions and each region has its own culture, i.e. language, religion, arts, food, etc. Obviously, women in India wear a unique clothing with colorful silk.

2.2.3 Research works on the cultural differences

Research on cultural differences has been widely studied in education [53][54], m-commerce [55][56], and communication [57]. These studies indicate the different cultures of each country. However, studying how a user's culture affects their usage of OSNs is a challenging task. For privacy concerns, Tsoi et al. [24] indicated that culture has an effect on Hong Kong and French users' behavior in OSNs. Hong Kong users are more likely to disclose personal information to others and make a connection with new users. Meanwhile, French users feel less in control when updating personal information and post only general information. To show the cultural effect of true commitment, Vasalou et al. [58] studied the cultural differences between US, UK, Italian, Greek, and French users of Facebook. The US and UK users give priority to groups, while Italian users rate groups, games, and applications as being the most important. As for the size of networks, Kim et al. [59] found that culture has a great influence on relationship maintenance. The US students make less effort in taking care of their relationships, whereas the Korean students tend to get social support from existing relationships. Ratikan et al. [60] analyzed the features and factors influencing users when consuming information. Furthermore, they studied the cultural differences between Japan and Thailand. They believe there are two things which could solve information overload and cultural ignorance problems.

However, as mentioned above, few previous research works have presented the cultural differences of readers when consuming information in OSNs. Therefore, studying this aspect is necessary in order to understand the way of thinking, feeling, and acting, when users in different countries consume information.

2.3 Information Feeding Mechanism (IFM) in OSNs

The IFM is an important part of OSNs. It has duty to select and display the information to the user. The amount and quality of information depends on the IFM.

2.3.1 Existing IFMs

- **Facebook** uses an EdgeRank algorithm [25]. It optimizes the user's news feed by using scoring to decide what information should appear. It considers the item (photo, text, video, etc.) displayed on the user's news feed as an object. When the object interacts with other users, such as in commenting, tagging photos, etc., the EdgeRank algorithm creates an edge. The edge has three components: affinity, weight, and time decay. Affinity represents the user's relationship with owner of an item and is assigned by score. When the user frequently connects to the owner of an item, the user will achieve a higher affinity score. Meanwhile, if the user has not interacted with his/her friend for 1-2 years, the affinity score will be very low. Weight refers to the user's actions. Interaction with videos, photos, and links are calculated as having the highest weight. Time decay indicates how recent the item is. New items have more chance of appearing on the news feed. Recently, Facebook has provided a 'hide' feature to the user. This feature allows the user to hide the information that he/she does not wish to see or disturbs the user, as shown in Figure 2.6. Facebook asks the user why he/she is hiding this information, as depicted in Figure 2.7.

- **Google+** [26] allows the user to see information from members in any circle via his/her Home Stream. Home Stream shows information coming from specific or public sharing, according to time. However, the user can filter the information by selecting specific circles. The user can control the amount of information appearing on the Home Stream by adjusting the volume (more, standard, and fewer) in each circle, what's hot and communities. This indicates that Google+ no longer shows all information in the user's Home Stream.

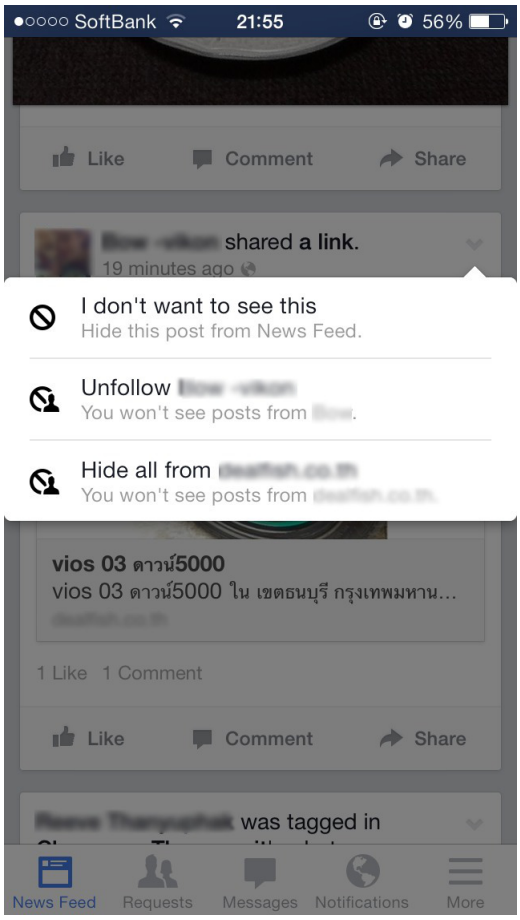


Figure 2.6: The hide feature in Facebook

- **Twitter** [39][40] uses the concept of social networking and SMS messaging. It indicates a simple way to provide the status of others via Twitter Timeline, where it shows tweets by using reverse chronological order. This means a new tweet is added to the front. The user can see the mix of tweets in the Timeline from following the people involved. Although Twitter allows the user to see all of the tweets on his/her Timeline, it provides a search engine to filter out unwanted or unrelated tweets and to discover new tweets.

However, existing IFMs do not emphasize the user's current situation, preference, and culture when serving the information via the SNP. Thus they can lead to information overload.

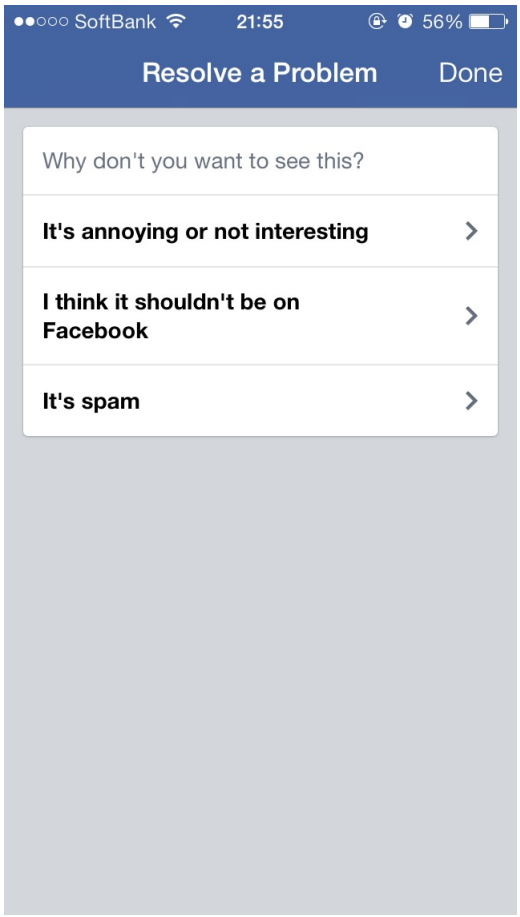


Figure 2.7: The hide feature in Facebook

2.3.2 Information overload

With the high performance of the Internet, creators can share thoughts, opinions, and feelings with other users via OSNs, i.e. Facebook, Google+, Twitter, etc., thus becoming authors and publishers of the information. Undoubtedly, information overload can cause a problem for readers, who seek interesting information on the SNP. This is because they rely on the quality and quantity of information provided by OSNs [61]. Information overload is defined by BusinessDictionary.com [62] cited in [20] "stress induced by reception of more information than is necessary to make a decision (or that can be understood and digested in the time available) and by attempts to deal with it with outdated time management practices". The information overload problem [63][64][65] is caused by three things: the cultural pattern of text, growth in the number of users by social graph, and sharing huge amounts of information.

Firstly, the cultural pattern of text. The reader might be unfamiliar with the writing style, language, and complexity of content [20]. Each culture has a different pattern of text. When readers have different cultures to the creators, the readers have to cope with this unfamiliarity. Therefore, it is easy for the readers to experience information overload.

Secondly, the growth in the number of users by social graph, which is consistent with the third instance. If the number of users increases, there is a high possibility that the creator will post information. Hence, the reader's SNP contains information from various sources, such as friends, family, and acquaintances.

Thirdly, sharing a huge amount of information. Generally, the creator has the freedom to post any type of information to his/her space provided by OSNs, e.g. text, photo, video, link, for any purposes anywhere at any time.

In order to show the consequences of information, Jones et al. [66] studied the effect of the number of messages on the readers. They found that if the readers often face information overload, they tend to respond to less complex messages, be unable to produce complex messages, and refrain from active participation. Moreover, many research works indicate that information overload causes decreased performance in activity. Zeldes et al. [67] found that this problem leads to a reduction in thinking, generating creative ideas and ways to solve problems. Schick et al. presented that consequences of the information overload problem can cause the readers to feel confused, anxious, and dissatisfied [17]. Bontcheva et al. [18] found that over 50% of Twitter users need some mechanism to filter irrelevant information. This emotion can lead to activity reduction in OSNs [19].

2.3.3 Information filtering

Information filtering is the process of reducing unimportant or redundant information to the users [68]. In the large community of social networking, techniques are needed to filter some information before displaying to the user. Therefore, the results of filtering might be more consistent with the user's interest. Using information filtering is considered to

be efficient for reducing information overload. Currently, three approaches are used for filtering information: content-based approach, collaborative or social filtering, and hybrid collaborative filtering [6][68].

1. **Content-based approach**

   Information is filtered by analyzing the user's profile using item description and matching regularities in the content. This approach needs sufficient detail from the user's profile, obtained either explicitly by interview or questionnaire, or implicitly from observed behavior, such as browsing features of the words or pages, and retrieving items the user liked in the past [69]. In this approach, learning algorithms for the data mining task are required, such as Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine (SVM). However, the cold start problem [6] can occur in the content-based approach, it needs to have enough data in the database to build the relevant information for the user. Previous works [70] in the content-based approach filter unwanted information by using a data mining algorithm with features, such as the user's profile, time schedule, and interest. Nakamura et al. [70] took into account the user's preference and timetable to block content that may spoil a user's enjoyment, while Loeb et al. [71] handled information overload and missing information by using specific contexts (location and time), mood, and social as features. Vanetti et al. [28] tried to filter unwanted information in OSNs by using a flexible rule-based approach as well as machine learning by automatically categorizing the content of information as a key component. They helped the user to restrict the information on his or her Social Network Page (SNP). Koroleva et al. [64] tried to solve information overload in OSNs by using a Neural Network algorithm to filter out irrelevant information together with features which were not provided by OSNs. The information is sorted by its level of importance.

2. **Collaborative filtering (CF) approach or social filtering approach**

   The recommender system is an active information filtering system that tries to predict information depending on the user's preference [68], and is widely adopted. For example, this approach is used by online shopping companies, such as Amazon,

eBay, Netfix, and so on in order to recommend new books or movies to their users, because it is easy to implement and has a high performance level [72]. Figure 2.9 indicates the result of the recommendations after using the CF approach.

This approach has three categories: memory-based CF, model-based CF, and Hybrid recommenders. Table 2.2 describes the techniques used in each category, with advantages and disadvantages. However, the concept of this approach uses information rated by other people who predict the same interests to a user. Therefore, this technique filters out inconsistent information for the user. It requires a large group of users to share the same preferences as the active users, and the system then makes recommendations of unknown preferences to the user [6]. For example, in Figure 2.8, all users 1, 2, and 3 like items U and V. Both users 2 and 3 have the same interest in items X, Y, and Z. Therefore, the system recommends items X, Y, and Z to user 1. The system refers to the user's preference by reference to the item he or she has rated. The rated items are recorded in a user-item rating matrix as shown in Table 2.3. Nonetheless, the cold start problem (some researchers call this problem a new user problem or a new item problem [73][74]) also occurs, meaning that a new user cannot be recommended to a new item because no other user has previously rated certain items before processing. This creates missing values in a matrix. Resnick et al. [75] applied the CF approach early on for recommendations in the user rating data to calculate a similarity value, making recommendations to those who seek advice. Changchun et al. [5] improved the algorithm for personalized recommendation by using the CF approach.

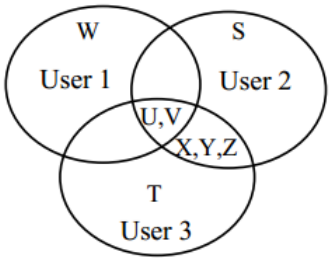


Figure 2.8: The working principles of collaborative filtering [5]

Table 2.2: Overview of CF approach

CF categories	Representative techniques	Main advantages	Main short comings
Memory-based CF	(a) Neighbor-based CF (item-based/user-based CF algorithm with Pearson/vector cosine correlation) (b) Item-based/user-based top-N recommendations	(a) easy implementation (b) new data can be added easily and incrementally (c) need not consider the content of the items being recommended (d) scale well with co-rated items	(a) are dependent on human ratings (b) performance decreases when data are sparse (c) cannot be recommended for new users and items (d) has limited scalability for large datasets
Model-based CF	(a) Bayesian belief nets CF (b) clustering CF (c) MDP-based CF (d) latent semantic CF (e) sparse factor analysis (f) CF using dimensionality reduction techniques, for example, SVD, PCA	(a) better addresses sparsity, scalability and other problems (b) improves prediction performance (c) finds an intuitive rationale for recommendations	(a) expensive model-building (b) has a trade-off between prediction performance and scalability (c) loses useful information for dimensionality reduction techniques
Hybrid recommenders	(a) content-based CF recommender, for example, Fab (b) content-boosted CF (c) hybrid CF combining memory-based and model-based CF algorithms, for example, Personality Diagnosis	(a) overcomes limitations of CF and content-based or other recommenders (b) improves prediction performance (c) overcome CF problems, such as sparsity and gray sheep	(a) has increased complexity and expense for implementation expense (b) needs external information not usually available

Table 2.3: An example of the user-item rating matrix [6]

	Shrek	**Snow White**	**Spider-man**	**Superman**
Alice	Like	Like		Like
Bob		Like	Dislike	Like
Chris		Dislike	Like	
Tony	Like		Dislike	?

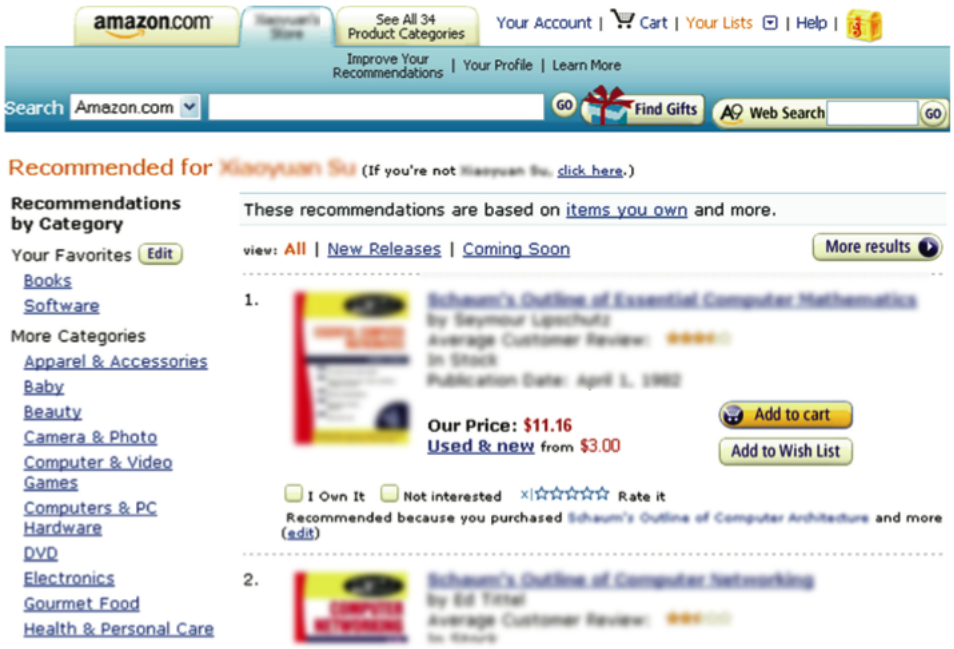


Figure 2.9: Use of the collaborative filtering system for product recommendation by Amazon [6]

3. **Hybrid collaborative filtering approach**

   This approach is a combination of the CF approach and other information filtering approaches, such as the content-based approach, or association of two classes of collaboration; memory-based and model-based. The object of this approach is to abstain from the limitation of both approaches described above and to improve the accuracy of prediction [6]. This removes unwanted information and recommends interesting information to the user. Prior works found [6][76] that hybrid CF, combining both the memory-based and model-based approach, gives more ac-

curate results than a combination of the CF approach and content-based approach. Moreover, it can address the new user and new item problem where the CF approach cannot achieve a satisfactory result due to insufficient data. Hannon et al. [77] focused on how information can help users to take advantage of Twitter. They use the content-based and the CF approach for recommendations. Claypool et al. [78] presented the hybrid CF approach by considering the weight of each user in content-based and CF predictions in order to solve the gray-sheep problem, where users do not rate agreement or disagreement for any group of people. This can result in the failure of the users to receive benefit from the CF approach. Miller et al. [79] proposed a way of protecting the user's privacy by a CF recommendation system.

These three approaches have different advantages and disadvantages. The CF approach and the hybrid CF approach require the collection of a large amount of ratings for the database, although the hybrid CF approach can solve the cold start problem [6]. Sufficient data needs to be obtained from the voting or sharing of users which is not an easy task. The CF approach relies on data from ratings, but we cannot know how reliable that data is, thus accuracy will be compromised and not good enough for filtering. For the content-based approach, existing research studies [70][71] remove inconsistent information by applying a few features, which may not be enough. However, the reason for this research is not to filter information dangerous to the user (referring to the reader) in the sense of protecting the reader's computer, but it attempts to serve relevant information on the reader's SNP according to the reader's current situation and preference.

2.4 Privacy protection models in OSNs

2.4.1 Privacy concerns

Privacy is a basic human right where an individual or group has the ability to decide what information about herself/himself/themselves is revealed to others [80]. When Web 2.0 was introduced, Internet privacy concerns were growing. This is because Web 2.0 presently allows the user to be more sociable in OSNs, e.g. Facebook, Google+, Twitter than in the past. Information sharing might not be suitable for making the information

available to all users. Nowadays, the creator tends to reveal personal information or collaborative information to the public by sharing space provided by OSNs [81], such as with News Feed in Facebook or Timeline in Twitter. Personal information refers to identifiable information, e.g. e-mail, photo, phone number, or schedule. The collaborative information indicates that the information does not belong to only one creator, but is associated with multiple users. For example, a photo can be taken by three creators, making it possible that a couple of them may lose their privacy unintentionally. Furthermore, if the information leaks to unwanted target readers (referring to those with whom the creator is not willing to share), it is hard to solve. One person cannot compel another to share. In other words, when the information has already been consumed by others, the creator cannot command those readers to stop spreading the information via mobile phone, word of mouth, and so on. Generally, the creator wishes to disclose information to small groups, such as friends and family [82]. Nonetheless, in some cases, the creator intends to reveal personal information to those readers who do not have a direct relationship with the creator [83][84]. The loss of privacy in OSNs can be caused in three ways:

Firstly, the co-owners do not have permission to control the collaborative information they are associated with. In addition, they might not realize that their information is being managed by others. The owners can normally tag, mention, or share the collaborative information with each other without asking permission. These actions can link directly to the profile of the creator [13]. However, OSNs enable anyone to visit another's profile if they do not use privacy settings. Figure 2.11 depicts the drinking of alcohol. It is possible that this photo is seen by someone in the family. Figure 2.10 shows a Rutgers University freshman posting a video of his roommate without permission. This creates a loss of that person's privacy. He may be ashamed of his actions and the fact that he took a wrong decision. In the case of Facebook, when the co-owners are tagged in a photo, other readers not associated with it, can observe the activity from his/her News Feed. For example, from user A's News Feed, they can know of user B's activity from: "user B was tagged by user C". Although the owner has rights to control publication of this photo, the co-owners should be asked for permission since they also have portrait rights. Therefore, the information might leak to unwanted target readers (with whom the owner

or the co-owners are not willing to share).

Secondly, sensitive information levels rely on cultural differences. In other words, seeing the same information can lead to different privacy concerns. For example, an owner from a Western country views a photo showing kissing in public as nothing special but the co-owners from Asian countries may think it is shameful and should not be spread on OSNs. Moreover, each culture has a different style of information sharing. For instance, Hong Kong creators are more likely to disclose personal information while French creators feel less in control when updating personal information [24].

Lastly, negotiation of privacy based on culture is a difficult task because the real meaning for the owner and the co-owners of different cultures needs to be understood. This negotiation is not carried out via face-to-face communication, and hence when the co-owners and owner are in different places, the owner cannot guess the real meaning from any clues in the co-owners' expressions.

Figure 2.10: An example of information leakage [7]

Figure 2.11: An example of information tagged by other users

2.4.2 Access control models

In order to protect the user's privacy, most research works have proposed an access control mechanism. Carminati et al. [85] proposed a rule-based access control mechanism for OSNs. Type, depth, and trust level of existing relationships between users were applied to express the complex privacy policy. Gollu et al. [86] presented a social-networking-based access control mechanism for information sharing. Identities of users were viewed as key pairing and social relationships. They provided a control list to determine who can access the information. Hart et al. [87] used the relationship information existing in OSNs for a content-based access control model. This model could authenticate the user to access the information. Ellison et al. [88] discovered the interesting analysis that the user tends to disclose information to online friends because they do not have contact with each other in the real-world. Hu et al. [13] proposed a mechanism for detecting and resolving privacy conflict among users with shared ownership of the collaborative information. Their research works enables these users to provide the policy then calculates the privacy risk and sharing loss. Hu et al. [89] presents collaborative privacy management for shared data in Google+, introducing the concept of circle and trust to their model. Squicciarini et al. [23] considered that information might not belong to only one user in some cases, therefore they created a mechanism to support information sharing in OSNs based on the notion of content ownership. They implemented a prototype system hosted in Facebook.

2.4.3 Other solutions for privacy protection

Besides the access control models for privacy protection, there are other solutions. Dinh et al. [90] attempted to construct a circle of trust by proposing the hybrid algorithm to investigate the maximum circle of trust problem. Thus, the user can safely share information with others, and it will not be leaked to unwanted target users. Li et al. [91] used machine learning techniques and structured semantic knowledge to gain knowledge of users' profiles and past privacy setting patterns, making recommendations for privacy settings to the users. Adu-Oppong et al. [92] applied automatically extracted network communities to make privacy policies easier by grouping friends into lists.

Although many access control models and other solutions have been proposed for privacy protection, they allow only the owner to control privacy management. Few research works realize the loss of privacy for co-owners associated with the collaborative information. In some research works [13][89], the owner and co-owners can create the privacy policy, but not everyone will be satisfied. The possibility of privacy violation remains if at least one co-owner intends to keep their information private.

2.5 Conclusion

This chapter starts with an explanation of the background and related works in OSNs such as Facebook, Google+, LinkedIn, and Twitter. OSNs have a major impact on communication because they enable the user to publish any information, e.g. personal information, photos, videos, and opinions to other users and to meet other users for various purposes, e.g. business, entertainment, or social. They are currently used by many people around the world, with Facebook having over one billion active users. Therefore, the study of cultural differences in IFMs of OSNs and privacy protection models are attracting growing attention and contributing to research works.

For the cultural differences section, definitions of backgrounds and national culture are

provided. However, there is no official culture definition since it depends on the context where the definition is applied. This research follows the definition: Culture is shared mental software, "the collective programming of the mind that distinguishes the members of one group or category of people from another" given by Hofstede [46]. This section explains why a nation can be used as the unit of analysis for cultural differences. In addition, related works concerning cultural differences in OSNs have shown that users in different countries each behave individually.

For IFMs in the OSN sections, the details of existing IFMs, EdgeRank of Facebook, Timeline of Google+, and Twitter are introduced. The existing IFMs can lead to an information overload problem (feeding too much information) because they do not realize the reader's cultural differences. This section also explains the information filtering techniques, content-based approach, CF approach, and hybrid CF approach, which can filter uninteresting information.

The privacy protection models in the OSN sections indicate privacy concerns and their causes. Literature reviews are then presented to try to solve the problem of privacy concerns by proposing access control models and other solutions.

Chapter 3

Research methodology

3.1 Introduction

The goal of this chapter is to introduce two main architectures for information feeding and privacy protection in Online Social Networks (OSNs). A user of OSNs can generally carry out two main processes as depicted in Figure 3.1.

These two processes provide the user with communication and social interaction in aspects such as business, entertainment, and politics. However, when the number of OSN users with different cultures increases, the volume, variety, and sensitivity levels of information also increase. Thus, the challenging task of this research is to reduce two main problems of OSNs: information overload and loss of privacy. This research covers information consumption and information sharing. The user can have the ability to consume relevant information with less annoyance based on the user's culture. The user can additionally share information such as feelings, opinions, status and so on with other users.

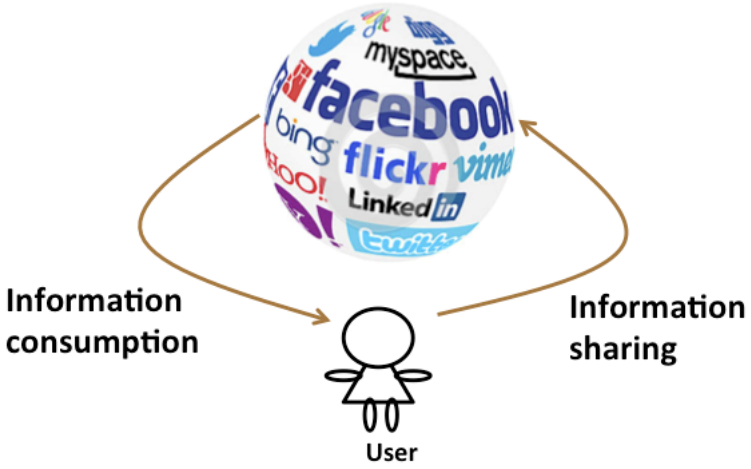


Figure 3.1: The user's abilities in OSNs (taken from www.dreamstime.com)

3.2 Architecture for information feeding and privacy protection by considering cultures in OSNs

3.2.1 information consumption

Generally, a user in OSNs can consume any information, such as text, photo, video, or link via the SNP. The user in this context is known as a reader. The reader faces information overload and cultural ignorance problems, when too much information appears and is not consistent with the reader's culture.

To overcome these problems, a new type of Information Feeding Mechanism (IFM) is proposed. Figure 3.2 displays an overview of system architecture for the proposed IFM. In this IFM, dynamically feeding interesting and important information is achieved by considering the reader's current situation and nationality. The proposed IFM consists of eight components: data collection, data analysis, a set of influential features and factors, a repository training set, data aggregation, OSNs, information filtering, and information organization. The core of the proposed IFM is the information filtering component. This component obtains useful information by data collection, social graph generation, reader's Profile, and feature extraction. Thereafter, the information organization component orders that information which is allowed to show on the Social Network Page (SNP)

according to reader's preference. Explanation of the eight components is given below.

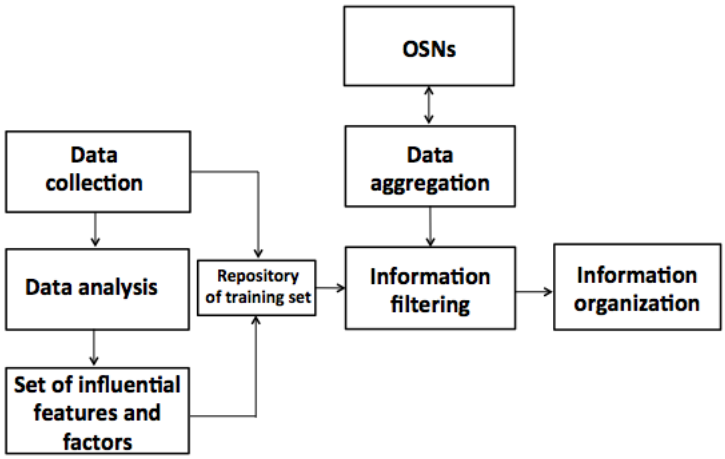


Figure 3.2: The proposed IFM for information consumption

1. **Data collection** is placed at the beginning when gathering data from the respondents. The set of features (e.g. the reader's current situation, information category, and information popularity) is set and defined before the start of the survey. Meanwhile, the set of factors (e.g. career, age, and gender) represents the respondent's data obtained from the survey. A questionnaire is designed for collecting the data together by studying the cultural differences between Japanese and Thai readers. It simulates several scenarios using a set of features to indicate characteristics from information on the reader's SNP.

2. **Data analysis** is used to prepare a set of influential features and factors to study the cultural differences in OSNs. The features and factors that have an impact on the reader's decision in selecting the information to consume are also analyzed. The data in the data collection component is investigated for cultural differences in certain aspects based on the set of influential features and factors.

3. **Set of influential features and factors** is the results obtained from the data analysis component. These features and factors are considered as important for the information filtering component as will be explained later. More details of influential features and factors will be shown in Chapter 4

4. **Repository of training set** is built using data from the data collection component, and the set of influential features and factors. This repository consists of two databases for Japan and Thailand used in the information filtering component.

5. **Data aggregation** stands for getting the amount of information that is being fed into the reader's SNP, including certain properties from OSNs, such as the creator's name, created time, and so. To achieve this, the component must ask permission from the reader before retrieving information. When this component has received such permission, it will request OSNs to obtain the information.

6. **OSNs** provide an application programming interface (API) for accessing the databases. Each OSN's API is different. For example, on Facebook, it allows only the sections having the required permission, to be retrieved, otherwise it cannot allow.

7. **Information filtering** is a core component of the proposed IFM. It tries to filter useless information by using a classification concept together with a set of influential features and factors. This research compares three classification algorithms: Decision Tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB) in order to find the best performance. The NB algorithm obtains the highest accuracy and the fastest processing speed. Therefore, when new information arrives in the SNP, it will be classified by the NB algorithm together with different sets of features and factors.

8. **Information organization** is introduced for ordering information by using the reader's preference. This differs from existing IFMs, which order posts by reverse chronological order or by top story category.

The work-flow of the proposed IFM starts from data collection, which requires a large amount of data from the survey for Japanese and Thais. This data is analyzed in order to compare the cultural differences in information consumption. In analysis, the data investigates the cultural differences based on the set of features, however the results of analysis are not very clear. Therefore, a set of factors is added during analysis. Then, both sets are investigated to find which features and factors have influenced the reader. Next, a set of influential features and factors of each country and the data in the data collection

component are collected in a training set repository, which is used as an input information filtering component. Subsequently, when new information arrives in the proposed IFM, it is classified by the NB algorithm in the information filtering component as to whether or not this information should be allowed to show on the SNP.

3.2.2 Information sharing

In OSNs, the user, known as the creator, can generate and publish any information for other users anywhere at any time. However, the creator is concerned with loss of privacy because he or she lacks adequate privacy protection. Thus, loss of privacy is a crucial problem in OSNs. Normally, when the creator posts information into OSNs, it is frequently viewed as being owned by its creator, or owner. However, in the real world, collaborative information (e.g. text, photo, video, or link) might simply not belong to the owner in some cases. It may be associated with multiple users, known as co-owners. Hence, the collaborative information might leak to unwanted target readers, resulting in loss of privacy for the owner and co-owners.

To address this problem, this research proposes collective privacy protection (CPP) to balance the need for privacy protection and information sharing as demonstrated in Figure 3.3. This enables the owner and co-owners to participate in a privacy policy and hold a majority vote over the collaborative information. It can identify and reduce privacy conflicts because at least one co-owner intends to remain private. This research aims to protect the privacy of the owner and co-owners. The collaborative information will not leak to unwanted target readers due to implementation of a maximum boundary. Furthermore, when the co-owners want to share information, he/she also can portray themselves as the owner and create the privacy policy. Five components of the proposed CPP are introduced.

1. **Social graph** creates a graph representing social relationships between the user and members in a contact list. This includes the relationship type or group, affinity levels between the user and members, and member preferences

2. **Privacy policy** is designed to limit the number of readers who can see the shared

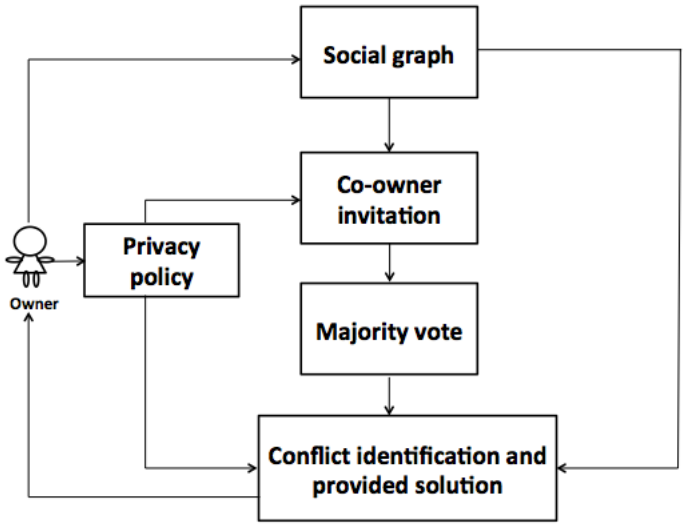


Figure 3.3: The proposed collective privacy protection for the owner and co-owners when information sharing

information. The idea of this policy is that the owner tries to match the shared information to the target readers who might be interested in it. The owner constructs a privacy policy by using four pieces of useful information from a social graph: type or group of relationship, distance of information propagation, affinity between the owner and target readers, and preference of target readers.

3. **Co-owner invitation** is proposed to inform co-owners that they are part of the shared information because they might generally be unaware that their information is being managed by others. This component is important since it helps co-owners acknowledge and vote on the privacy policy created by the owner.

4. **Majority vote** involves a duty to seek the consent of all co-owners as much as possible because allowing all owner and co-owners to create a privacy policy is difficult to please everyone. This component starts with gathering information of the status of all co-owners using the co-owner invitation component. Nonetheless, it will take time to collect voting responses from all co-owners; so it needs a specific time-frame. The co-owner with no response status will be moved to rejection status to protect privacy when the time limit is reached. The advantage of majority vote is that it still provides privacy protection for those co-owners who reject the privacy

policy.

5. **Conflict identification and provided solution** is designed to find conflicts among co-owners who accept and reject the privacy policy and create a solution for those conflicts. This component finds the cause of conflict to provide a suitable solution, and a list of target readers is recommended to the owner. The owner can verify the information before posting it.

The work-flow of collective privacy protection begins with the owner. The co-owners are detected in order to send an invitation that he/she owns part of the information and to inform co-owners that their information is being posted to the owner in OSNs. When co-owners receive the invitation and read the privacy policy, he/she can vote on it. The co-owner can use one of three statuses: acceptance, rejection, and no response. Acceptance means the co-owner agrees with the privacy policy. Rejection indicates the co-owner declines the privacy policy or he/she needs to keep their information private. The no response status indicates that the co-owner did not accept or reject in time. Nevertheless, when the time limit is reached, those co-owners with no response will change their status to rejection because privacy is considered as a high priority. Next, the proposed CPP detects the privacy conflict between the owner and co-owners, and provides a solution for each conflict. A list of the target readers who can see this information based on the privacy policy is suggested to the owner for them to re-check before uploading it on to the OSN.

3.3 Contribution of architectures

According to the two architectures; a proposed IFM for information consumption and a proposed CPP using majority vote for information sharing. The contributions are shown below.

There are three main advantages to information consumption. Firstly, the proposed IFM filters uninteresting and unimportant information in order to reduce the amount of information on the SNP using the NB algorithm together with influential features and

factors.

Secondly, the reader can dynamically consume interesting and important information in a short period of time based on the current situation. This information is ordered by reader preference.

Thirdly, the proposed IFM can decrease the limitation of traditional IFMs, such as information overload, lack of information, and providing uninteresting and unimportant information. It can also be applied to other cultures, societies, and businesses.

The usefulness of the proposed CPP by using majority vote is explained below. Firstly, the privacy policy can limit the number of readers who can see the information. Using the privacy policy will let the owner know who can see the information more efficiently. Creating the privacy policy is flexible for the purpose of sharing information. For example, if the owner needs help from others, he/she wants information to be spread quickly for as long as possible. Another case is that if the owner shares information only with family, he/she can customize the privacy setting by selecting a type or group of relationships.

Secondly, the co-owner invitation is an important component of information sharing. This component lets the co-owners know that their information is being managed by others. In many cases, the co-owners lose privacy due to the owner sharing the collaborative information without permission. Thus, the co-owner invitation component can create different results when compared with existing research works. It can prevent co-owners from leaking information to unwanted target readers. For example, requesting permission from co-owners for sharing photo can avoid the portrait or photo rights violation.

Thirdly, the majority vote concept in the proposed CPP allows consent to follow the privacy policy where more than half have voted. However, the proposed CPP still supports co-owners who reject the privacy policy. As a result, the privacy of the co-owners with rejection status is protected although the overall vote result of the policy is accepted.

3.4 Conclusion

This chapter mainly explains two main architectures: the proposed IFM for information consumption and the proposed CPP by using majority vote for information sharing. These two architectures are necessary and important for OSNs because they help support the user in social interaction and communication. For information consumption, the proposed IFM consists of eight components: data collection, data analysis, a set of influential features and factors, a repository of training set, data aggregation, OSNs, information filtering, and information organization. The advantages of the proposed IFM are that it reduces the amount of information, serves interesting information, and decreases the limitation of existing IFMs. Moreover, it can be applied to other cultures, societies, and businesses.

For information sharing, the proposed CPP by using majority vote involves five components: social graph generation, privacy policy, co-owner invitation, majority vote, and conflict identification and solution. Its advantages are that it limits the number of readers, protecting the privacy of not only the owner but also the co-owners, giving the chance for co-owners to participate in the privacy policy created by the owner. Thus, this architecture can provide different results compared to other research works.

Further explanation of cultural analysis is presented in Chapter 4, indicating the preparation of features, questionnaires for the survey, as well as data analysis. Chapter 5 describes the use of results from Chapter 4 as training data for information filtering, experimental setup, and discussion of the results of Japanese and Thai respondents. In Chapter 6, the proposed CPP is presented by using majority vote, together with details of each component.

Chapter 4

The impact of cultural differences for information feeding

4.1 Introduction

There are many instances to indicate that people in different countries have different behavior, feelings, and ways of thinking. According to communication style, there are high context (HC) culture and low context (LC) culture. The concept of HC and LC cultures relates to implicit and explicit ways of communication [93]. The HC culture is considered as an indirect method of communication. People in this culture such as Japan, China, and Thailand emphasize interpersonal relationships. Words might not play an important role in communication. The listener is expected to look for context clues from the speaker's tone of voice, facial expression, gestures, and posture to understand what the speaker means. On the other hand, the LC culture as in America and Canada will know precisely what the speaker means by words alone. People in this culture try to carry a message by words rather than by nonverbal means. Verbal message is explicit.

Due to the differences in communication style, I think the Japanese and Thais might have different cultures in information consumption from Online Social Networks (OSNs), which nowadays play an important role in communication and social interaction. At present, Information Feeding Mechanism (IFM), which has a duty to select and display information to readers in OSNs, is developed without considering the cultural differences.

Readers in different countries have their own cultural behavior and criteria, when receiving the information to consume. Thus, the goals of this chapter are to prepare the influential features and factors and study the cultural differences in information consumption in OSNs. The analyzed results in this chapter are valuable for feeding information based on the reader's culture in OSNs since they can result in understanding each process of thoughts, feelings, actions, and so on. Furthermore, the analysis is beneficial for marketing strategies in online business. To do this end, several components are required as described in Section 4.3.

4.2 Country selection for studying the cultural differences

Many research works [94][95] indicate that Asian and Western countries have big differences, such as communication style, lifestyle, learning style, and so on. Thus, studying the cultural differences between Western and Asian countries is not challenging because we know already that Western and Asian people have these differences. However, studying and understanding the cultural differences of Asian countries is very interesting. For this research, Japan and Thailand are selected for investigation into cultural differences. Although both countries are located in Asia, I believe that Japanese and Thais do not have many shared ways of thinking, feelings, and behavior. This is because these countries have their own unique cultures, which have evolved from the past into the present.

Japan has a unique culture compared to other countries in the world and differs from the cultures of Asian countries such as Thailand (Southeast Asia), Afghanistan (South Asia), and China (East Asia). Japan has a multifaceted traditional culture, gained over thousands of years, such as lifestyle, art, and social convention. Although Japan has opened itself to international trade and exchanged cultures with other countries over time, it still preserves its own culture, such as traditional sports, bowing, and taking off footwear. One example of the uncertainty avoidance (UAI) dimension defined by Hofstede [41] explains that Japanese tend to suffer anxiety in unexpected situations. Therefore, they prepare themselves by creating plans and predictions for coping with those situa-

tions. Japanese do not readily accept change.

On the other hand, the people of *Thailand* are quite flexible and easily accept and mix other cultures with their own. Thailand has adopted Western cultures such as social drinking culture, entertainment industry, and language. Recently, Korean cultures and Japanese cultures have influenced Thai traditional culture in the entertainment industry, and many imported products. Nonetheless, Thais try to preserve their traditional culture, such as with the non-verbal indicator of the *wai*, the linguistic status indicators of *phi* (for an older person) and *nong* (for a younger person). Thais also avoid conflict and make an effort to interact in a pleasant manner. Therefore, Thais have humility and a relaxed attitude by continually smiling. Culturally, Thais feel more comfortable in unexpected situations than Japanese. To reduce the degree of uncertainty, Thais apply strict rules, laws, policies, and regulations for controlling unexpected situations.

Even though this research studies the cultural differences of Japanese and Thais, the results can be applied to other cultures, where cultural characteristics are similar. For example, Japan is located in Asia but the results from Japan might apply to France because these countries are similar in communication style (high context culture) and UAI dimension (high UAI score) as mentioned in Hofstede [41].

4.3 Architecture and methodology

To prepare a set of influential features and factors and to study the cultural differences, we require two main components: data collection and data analysis. Figure 4.1 indicates the work-flow of each component so as to achieve the relevant goals and starts with the *data collection* component. This component is required to define features that might impact on the reader's decision. These features are used for designing a questionnaire in order to indicate the characteristics of the information in the Social Network Page (SNP), such as the News Feed on Facebook and Timeline on Twitter. The designed questionnaire asks the respondents for general information and their experience in OSNs. After that, all the data is gathered and kept separately in the database for Japan and Thailand. Before using these two databases, the data must be cleaned so that the databases are qualified

for use.

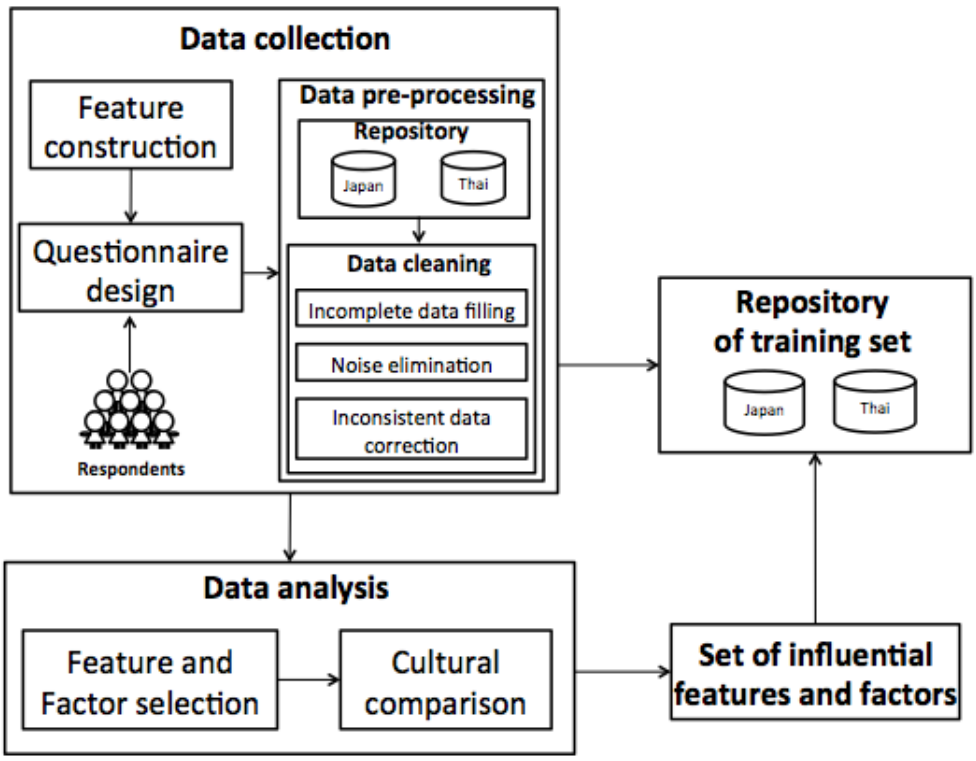


Figure 4.1: A study of cultural differences for information consumption in OSNs

The *data analysis* component attempts to select features which influence the readers' decision of whether or not the information should be fed into their SNP. However, the analysis results show that three out of seven features might not be sufficient for filtering the uninteresting information. The study shows that using factors arising from answers to the questionnaire impact on the readers' decision. Therefore, these factors are included during analysis. This component also compares the cultural similarities and differences between Japanese and Thais. More details of each component are described in the remaining sections.

4.4 Data collection

The objective of the data collection component is to gather data from the respondents, who have completed the questionnaire. The questionnaire is formulated using seven features

as explained in Section 4.4.1, in order to indicate the characteristics of information on the reader's SNP.

4.4.1 Feature construction

Previous studies have shown that almost all available data was used in features for filtering uninteresting information on OSNs [96][97]. Paek et al. [29] investigated a set of features to predict important information on News Feed by using around 50 features and using a machine learning algorithm. The features used in Paek's research concerned social media properties (i.e. the number of comments, the number pressing the like button, time decay, etc.), message text and corpus (i.e. the amount of words in a text, the number of stop words, ratio of stop words, total words, etc.), and shared background information (i.e. affiliation, religion, education, etc.). Based on Paek's features, the classification accuracy for important information on News Feed is about 64%. Chen et al. [96] studied three separate dimensions: content relevance, content sources, and social voting, for recommending interesting information to Twitter's readers. The best performing algorithm could achieve 72% of interesting information. Most studies [25][26][97][98] indicated that information category, affinity levels between the reader and creator, information popularity, and time decay were the most important features.

In previous works, especially Paek's research, many features are used for filtering uninteresting information in OSNs, but the classification accuracy is not very high. Some features can lose classification accuracy because they are not influential to the reader. In addition, some of them are not beneficial for solving the information overload problem because they might not be appropriately designed to take into account directly the cultural differences in OSNs as described in Chapter 1.

Figure 4.2 indicates the source of each feature for information filtering in the proposed IFM. This research focuses on reducing the information overload problem by considering cultures in OSNs, and follows the Hofstede dimensions, consisting of six dimensions as shown in Chapter 2. Five features are extracted using four out of six dimensions, namely the reader's current situation, type or group of relationship between reader and creator,

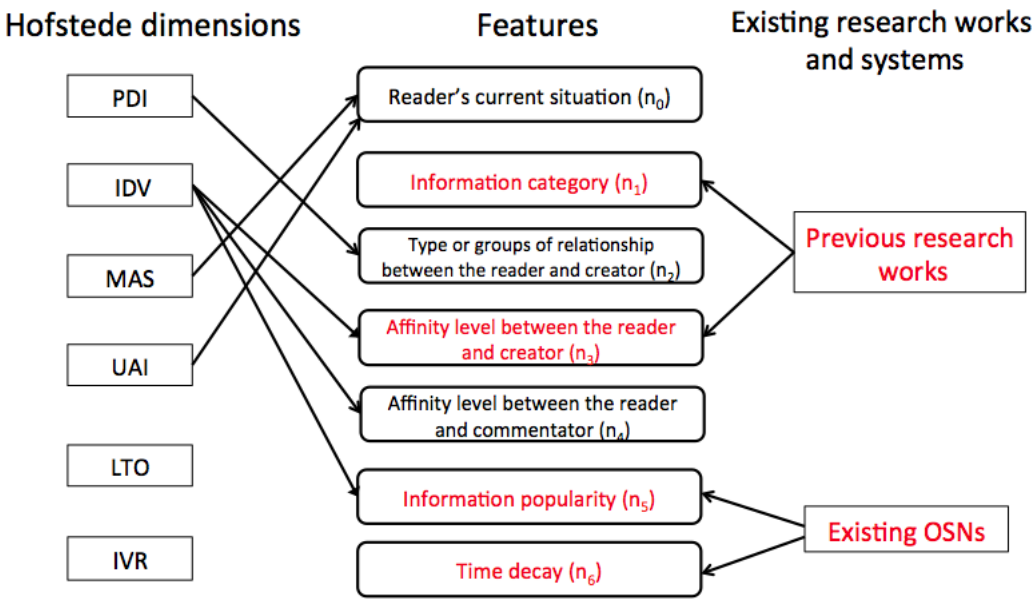


Figure 4.2: Feature construction for the proposed IFM

affinity level between reader and creator, affinity level between reader and commentator, and information popularity. These five features can indicate the characteristics of readers coming from different cultures. However, the LTO and IVR dimensions might not relate to communication in OSNs. The LTO dimension represents fostering virtues oriented towards the past, present, and future. The IVR dimension represents the degree to which a society allows relatively free gratification of basic and natural human drives concerned with enjoying life and having fun. Previous works [28][96][97] tend to use the information category feature and affinity level between reader and creator for filtering uninteresting information. In addition, the information popularity feature and the time decay feature were commonly used in existing OSNs, such as Google+, and Twitter. This study simulates several scenarios to indicate the characteristics of information in OSNs by the seven features described in Section 4.4.2. Further details of each feature are described as follows:

- **Reader's current situation (n_0)**

  This feature indicates the activity being performed by the reader in each period of the day, such as work, private time, shopping, travel, partying, meetings, and so on. Different situations influence the information category that the reader wishes to consume.

Moreover, readers in different cultures perform in different ways. This can be explained by the UAI and Masculinity versus Femininity (MAS) dimensions defined by Hofstede [41]. The UAI refers to a tolerance level when one faces an ambiguous situation or an employee expresses their intention to stay with the company for a long-term career. On the other hand, the MAS relates to the attitude of certain people in society towards achievement or nurture. Masculinity stands for a society focusing on achievement, assertiveness, and material rewards for success. Femininity stands for a society which emphasizes modesty, tenderness, and concern for quality of life. In countries with high UAI and MAS indices like Japan, Hungary, and Venezuela, employees have an emotional need to be busy and work hard. Also, they are motivated by competition to succeed for their team or organization. Therefore, employees in this group of countries pay attention to an assigned task and will stay late in the office to complete it. After work, they frequently enjoy a party to relax from the stress of their work. This may indicate that the behavior of people in low and high UAI and MAS indices are different.

The current situation can be reflected in many ways. For example, the reader directly inputs, or the system observes the reader's context from the Google calendar [99], Check-in feature of Facebook [10], and Twitter [12].

- **Information category ($\mathbf{n}_1$)**

  This refers to the information topic. Since OSNs allow the creator to post any information, they become sources of enormous amounts of information. Use of this feature is necessary for information filtering and is associated with reader preference. Moreover, if the content of an information category matches current circumstances in the reader's country, and the reader is interested in this information, the relevant category helps the reader to save time in consuming interesting in formation on the SNP. The circumstances of each country might be different and changeable. For example, the current circumstances in Thailand relate to politics due to conflict between the government and demonstrators. Therefore, it is possible that Thai readers are interested in the political information category. On the other hand, the infor-

mation categories relating to sport, especially football might have attracted Brazilian readers when the 2014 FIFA World Cup was being held in Brazil. Therefore, matching the information category to preference will increase interest for readers. The information is classified into categories, such as sport, music, advertisement, travel, personal stories, work, and games.

- **Type or groups of relationship between the reader and creator (n_2)**

  In OSNs, the reader usually has relationships with the creator of the information, such as employer, friends, family, and so on. This relationship sometimes indicates how important the information is. For example, if the creator is the reader's employer, this information might be a project, task, meeting, or something else of importance. This is because each society is unequal, and hierarchical as described in the power distance dimension by Hofstede [41]. People respect groups of people who are at the top level of hierarchy and respect for older persons is considered virtue. For example, in a company, the employer or manager has the power to make decisions. In addition, subordinates show loyalty, respect, and deference to their superiors. In a family, children are expected to be obedient to their parents.

  Therefore, the reader makes a decision to read the information by considering relationships or positions in society. Using this feature differs from existing OSNs, where every member is a friend.

- **Affinity level between the reader and creator (n_3)**

  This shows the familiarity between the reader and creator of the information or how often they interact [25]. The reader cannot generally apply the same affinity level to every member in a contact list. For instance, Reader A has 20 friends in a group from university. Reader A cannot treat or interact with every friend equally. In this research, three affinity levels between the reader and creator are defined as high, medium, and low. The high affinity level means the reader and creator interact frequently every day, such as commenting or pressing like (in Facebook) or +1 (Google+) button. The medium affinity level refers to those they usually interplay with every week or month. The low affinity level indicates that the reader has not

visited the page of the information owner for one year.

Based on the Individualism versus Collectivism (IDV) dimension introduced by Hofstede [41], this indicates a degree of interdependence on people in society. In most countries with a low IDV index (collectivism), people are generally expected to take care and support members of a particular group. In OSNs, the reader from collectivist countries, e.g. Thailand, Vietnam, and Indonesia maintains good relationships with other users in particular groups by interacting and consuming information posted by the creator coming from the same group with a high affinity level. The reader from individualist countries, e.g. America, Australia, and Great Britain might consume information without considering affinity levels if the information is interesting to him/her.

- **Affinity level between the reader and commentator ($\mathbf{n_4}$)**

  This feature represents the closeness of the relationship between the reader and commentator and can be considered when two pieces of information are created by the same creator and achieves a similar popularity level but with different groups of commentators. For example, there are two pieces of information appearing on Alice's SNP: A and B. Both are created by Bob, who is a friend of Alice. These two pieces of information obtained 100 comments and 50 like button presses. Nonetheless, they have different groups of commentators. Information A obtained comments and like button presses from most friends with whom Alice is familiar. On the other hand information B was received from Bob's co-workers and boss, who are acquaintances of Alice. Therefore, Alice can make a decision by using this affinity level, as to which information should be selected. This feature also has three affinity levels: high, medium, and low, and an explanation how this influences culture is described in the n_3 feature.

- **Information popularity ($\mathbf{n_5}$)**

  This refers to top stories currently of interest to people. Regularly, when the commentator presses the like (in Facebook) or +1 (Google+) button, watches a video, or interaction with other commentators to this information, the reader can observe

its popularity. Therefore, it is possible that the reader might be brought up to date by reading this information. Videos, photos, and links are generally considered to have the highest weight [25].

According to the IDV dimension, preventing loss of face in collectivist countries is also sensitive, and therefore the members of the group will help to support each other. In OSNs, when the creator in a collectivist country posts information, the reader tends to interact in order to maintain the relationship. As a result, this information becomes popular and is interesting to other readers. This differs from individualist countries where the reader's individual interest is expected to prevail over collective interest [41]. The readers from the countries with a high IDV index might not be interested in popular information.

- **Time decay(n_6)**

  Time decay shows how up to date the information is [25]. If the information has just been created, it tends to be interesting to the reader due to its newness. On the other hand, information that has been created for a long time is more likely to be dropped out of the reader's SNP because of its obsoleteness.

  The importance of time is depends on each culture. In some cultures such as Japanese [100] and Muslim [101], time is considered more important than in Thai culture. For example, transportation in Japan is famous for its punctuality and high quality services, i.e. train, bus, mail service, etc. On the other hand, time can be flexible in some situations in Thai culture. Accordingly the importance of time when consuming information, up-to-date information [102][103] is a priority for Japanese. For example, this information helps them to know as soon as possible if a natural disaster will hit their country as the Japanese have recently faced tsunamis and earthquakes several times.

4.4.2 Questionnaire design

In the survey, two questionnaires for each country, Japan and Thailand, are provided to study the cultural differences. The content of these questionnaires is similar, but the language used is different. They are simulated under several scenarios to indicate the characteristics of information posted by creators in OSNs by seven features as explained in Section 4.4.1.

The respondents numbered 161, consisting of 51 Japanese and 110 Thais. All of them were asked about age, career, interests, frequency of OSN service usage, etc. Table 4.1 indicates that most Japanese respondents were aged between 23 and 29 years old and most of Thai respondents were aged between 26 and 30 years old. 55% of the respondents were female and 45% were male.

Table 4.1: Summary of personal information supplied by Japanese and Thai respondents

Demography	**Japan**	**Thailand**
Gender	21 males and 30 females	52 males and 58 females
Age	23-29 years old	26-30 years old
Career	47.9% Students, 24.7%IT specialists, 22.3% Engineers, 5.1% others	
Interest	Sport, Music, Travel, Games	
OSNs service used	Facebook, Google+, Twitter, Mixi, etc.	Facebook, Google+, Twitter
Frequency of OSNs service usage	Less than 30 minutes-2 hours per day	1-5 hours per day

Each respondent answered 45 questions and were required to imagine themselves in particular scenarios. The scenarios were randomly selected by the respondents. In this questionnaire there were many types of scenarios, coming from a combination of seven features. Each type of scenario had an equal chance to be selected, and an attempt was made to simulate each of them as much as possible to the respondent. The respondent then answered the question: "If you see information with different scenarios 10 times in the SNP, how many times do you review this information?" Multiple choices are supplied to indicate the number of times. An example scenario is shown below.

"*There is information about a* ***sport (team, e.g. football, volleyball, etc.)*** *fed*

into your SNP. The owner of this information is ***a friend from university*** *and you have* ***usually interacted with your friend every week or every month. 10-30 people comment or press the "like" button*** *for this information. You and another person commenting on this information have* ***regularly interacted every day****. This information has* ***just been created****".*

Next, each respondent was asked: "Will you allow this information to be fed into your SNP when you are meeting with 10 co-workers". The bold text means seven features. The respondent then takes a decision either to "Allow" or "NOT Allow". This scenario requires the respondent to imagine 10 times. This means that the respondent will face this kind of scenario with different content but still on the same topic. For example, the information category is a sports team. The respondent could imagine that the information might relate to football or any other kind of team sport; however, the main topic of the information remains the sport. The number of times the information is reviewed by the respondents (No.Times) is used to define the degree of boredom of the respondent when he/she consumes the same kind of information several times. 0-1 times presents that if the respondent sees this kind of information 10 times, he/she will review it only once or not at all. This means that the respondent is extremely bored or dislikes this kind of information.

Seven topics are used in the experiment because they commonly occur in OSNs.

- **Personal stories** describe personal experiences, updating of status, self-promotion, etc.
- **Advertisements** show that OSNs have been recently used by companies, members in a contact list for online business, part-time job, or promoting products at a discount price. For example, Facebook allows advertisements to appear because it believes [104] that "Everyone wants to know what their friends like".
- **Game** indicates information about a game invitation or an opinion about a game, such as Dragon City, Farm Ville, and so on.
- **Working** was used to explain tasks, schedule plans, etc.

- **Travel** presents a journey to a place such as somewhere famous, an remote place, a natural attraction, etc. Some owners tried to review a place where they have never been.

- **Sport** indicates information about football, basketball, volleyball, etc. Usually, most people give feedback or opinions to those sports after they have watched them.

- **Music** refers to sharing or posting interesting music, popular music, old music, etc.

4.4.3 Data pre-processing

The data collected from the questionnaire as explained in Section 4.4.2, was kept in separate databases for Japanese and Thais. In a real-world database, we need to ensure that the data is ready to use. If the data is of poor quality, the information filtering component cannot produce quality results. Therefore, data pre-processing is necessary.

1. **Repository**

   Two databases for Japan and Thailand were used for collecting all data from the survey. Figure 4.3 provides a description of the databases consisting of 19 fields: age, career, gender, and the respondent's decision to "Allow" or "NOT Allow". "Allow" means the respondent allows this kind of information to be fed into the SNP. "NOT Allow" means the respondent does not allow. MySQL is used for database management, such as creating a database, updating a database, and so on.

2. **Data cleaning**

   Normally, in the real-world database problems can occur with missing data (lacking values), noisy or inconsistent data (containing some errors or something that stands out as being different from the rest). To improve data quality, missing values need to be supplied, together with identifying noisy data and correcting inconsistent data.

   - *Missing values* might occur in several situations, such as equipment malfunction, or values not fulfilled thanks to misunderstandings. These are perhaps indicated by a dash or blank [105], and thus the missing values need to be

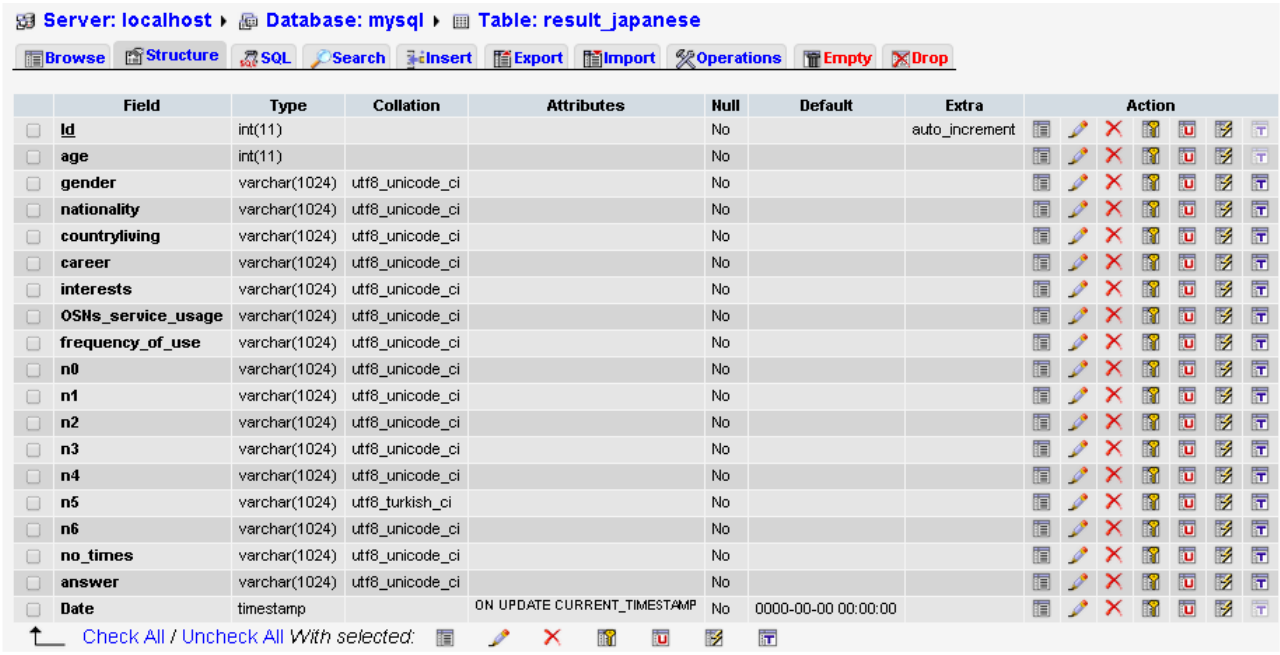


Figure 4.3: Description of design database

supplied. For instance, if some values are missed, as shown in Figure 4.4, calculation of the further component be incorrect and affect overall accuracy.

Name	Age	Gender	Career
Alice	24	Female	
Bob		Male	Student
Carol	30	Female	Teacher
David	28		Engineer
Eric	19		IT officer

Figure 4.4: Missing data in the database

- *Noisy and inconsistent data* is sometimes difficult to identify. It could be a random error or something that differs from the rest. This might be due to a data entry problem, data transmission problem, duplicate records, and human or computer error. For this reason, this kind of data should be detected and smoothed out. Figure 4.5 shows how the noisy data can be handled.

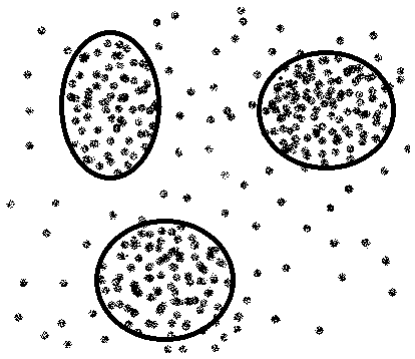


Figure 4.5: Noisy data handling by clustering analysis

4.5 Data analysis

The objective of data analysis is to prepare a set of influential features and factors and to study the cultural differences between Japanese and Thais in OSNs. Firstly, the feature selection tool in Waikato Environment for Knowledge Analysis (WEKA) [105] is used to investigate the hypothesis in Section 4.4.1. Similarities and differences between the Japanese and Thai cultures are then compared using several aspects based on the set of influential features and factors.

4.5.1 Feature and factor selection

Feature selection is used to select a subset of relevant features influencing the respondents' decision. Reducing uninteresting features in the database is necessary for the information filtering component. This is because a lot of features in the database not impacting on the respondents' decision will cause a drop in classification accuracy. In this research, features are investigated by using the feature selection tool from WEKA [105]. This research uses GainRatioAttributeEval as an attribute evaluator and Ranker as the search method shown in Figure 4.6. The GainRatioAttributeEval is a kind of a Single-Attribute Evaluator used together with the Ranker search method for generating a ranked list. The GainRatioAttributeEval measures the attribute by considering the gain ratio with respect to the class (the class is an answer of "Allow" or "NOT Allow" in this research).

Table 4.2 shows the influential value of each feature. This value indicates how the features impact on the respondents' decision whether or not to allow certain kinds of

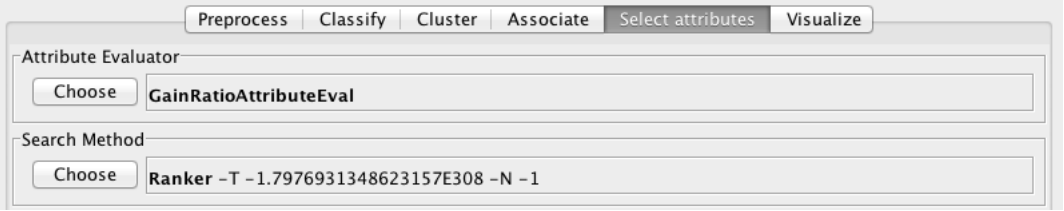


Figure 4.6: The attribute evaluator and search method used in the feature selection

information to be fed into their SNP. Japanese and Thai respondents place importance on each feature unequally.

Table 4.2: Influential features for Japan and Thailand

Japan		**Thailand**	
Feature	**Influential value**	**Feature**	**Influential value**
n_0	0.0124	n_1	0.0182
n_1	0.0114	n_0	0.0154
n_6	0.0011	n_2	0.0037

In this research, the top three rankings were considered. The first three influential features for Japanese respondents were n_0, n_1 and n_6 with influential values of 0.0124, 0.0114, and 0.0011, respectively. On the other hand, the n_1, n_0 and n_2 features with influential values of 0.0182, 0.0154, and 0.0037 serially impact on Thai respondents [60]. The most important feature of both countries was clearly different in that the n_0 and n_1 features were given precedence by Japanese and Thai respondents, respectively. This shows that most Japanese respondents emphasized n_0 or reader's current situation, when consuming the information in OSNs whereas most Thai respondents pointed to n_1 or information category as the most important. Even though the n_6 and n_2 features were ranked third for both countries, the influential values were quite a far distance from the first and second ranking. Other features (Japan: n_2, n_3, n_4 and n_5, Thai: n_3, n_4, n_5, n_6) not mentioned, had little impact on the respondent's decision. However, the behavior of both countries in is analyzed in the following sections.

Even if the first three features of each country were considered to impact on the respondents, their influential values were not very high. Quinn et al. [106] and Pfeil et al. [107] found that age impacts on behavioral differences between younger and older users in OSNs. Hargittai [108] presented that a person's demographic information, e.g. gender, race, and ethnicity is associated with use of OSNs. Therefore, the factors obtained from the questionnaire were applied to analyze behavior. For example, when the age of the respondents increases, Japanese and Thai respondents predominantly answered "Not Allow", especially Thai respondents older than 30 years old. All of the factors and answers from respondents are analyzed and investigated to find which factors have influenced the respondent's decision. The results as illustrated in Table 4.3 expose that the number of times information is reviewed by the respondents (No.Times) and age factors had a greater impact on their decisions. Table 4.4 summarizes the influential features and factors used in this research.

Based on the feature selection tool provided by WEKA [105], those features and factors having relatively little influence on the respondents' decision were removed. Classification accuracy in the information filtering component in Chapter 5 will increase because classification calculations are not disturbed by them. Furthermore, time consumption for calculation in classification will decrease since it uses only the data that impacts on the respondent's decision. In this sense, the influential features and factors are shown to benefit from the overall performance of the proposed IFM, as described in Chapter 5. Analysis of the cultural differences between Japan and Thailand will rely on the features and factors shown in Table 4.4. More details for each influential feature and factor are provided in Section 4.5.2.

4.5.2 Cultural comparison

1. **Influence of the reader's current situation (n_0) and information category (n_1) features on Japanese and Thai respondents' decision**

   Section 4.5.1 shows that the n_0 and n_1 feature influenced the Japanese and Thai respondents' decision on whether or not they allowed certain kinds of information to be fed into their SNP. The kind of information causing annoyance to the respon-

Table 4.3: Influential factors for Japanese and Thai

Japan		**Thailand**	
Factor	**Influential value**	**Factor**	**Influential value**
No.Times	0.1059	No.Times	0.0532
Age	0.0573	Age	0.0234
Preference	0.0168	Career	0.0102

Table 4.4: Influential features and factors for Japanese and Thai

Nationality	**Features**	**Factors**
Japanese	n_0, n_1, and n_6	The number of times a post is reviewed by the examinee (No.Times), age and preference
Thai	n_1, n_0, and n_2	No.Times, age and career

dents or that which should not be shown on the SNP when the respondents were in different situations was analyzed by using the n_0 and n_1 features.

Table 4.5 shows the results from analysis of the n_0 and n_1 features in great detail. Six respondents' current situations for the n_0 feature and information category (n_1) were analyzed, respectively. Other percentages not reported in Table 4.5, are not significant results. The results show that over 88% of Japanese respondents did not want to read advertisements (offering discount prices and part-time jobs) and games situations (meeting, working, or traveling). Moreover, studying Japanese respondents' opinions from interviews, some respondents stated that they often skip unwanted information. For example, if they were traveling, they would disregard an advertisement (about discount prices) immediately. However, if they received the feed during shopping, they felt this information was useful and increased their opportunity to buy something. Therefore, this group of respondents expressed that

if information could be shown according to the current situation, it might increase their interest level of that information.

For Thailand, two kinds of information were strongly uninteresting: advertisements (for part-time jobs) and games, especially during meetings. Clearly, around 90% of Thai respondents were not interested in advertisements about part-time jobs during meetings, work, private time, or when traveling. Also, around 80% of them ignored games during work time. During interviews with three Thai respondents as to whether the n_1 and n_0 features had an impact on their decision making and if they allowed the information to be fed into their SNP. Two of them reported that they contacted customers or co-workers via OSNs during meetings. Also, their boss assigned a task to a project team via OSNs like Facebook instead of e-mail; commonly used in their team projects. Moreover, it was found that showing the information at inappropriate time might cause distress to the respondents. For example, if they booked a holiday and then checked news updates or stories of others on Facebook during their trip, after seeing information about an assigned task they became worried about it. One interviewee said that she was bored with the information posted by her friends for self-promotion. Moreover, she thought that reading too much information via the SNP could make it difficult to find interesting information.

2. **Influence of the No.Times factor on Japanese and Thai respondents' decision**

   From the questionnaire, the respondents were asked: "If you see the information with different scenarios 10 times in the SNP, how many times do you review this information?" This is to investigate how the No.Times factor influences the respondents' decision.

   Figure 4.7 shows different graph patterns between Japanese and Thai. When the No.Times gradually increased, Japanese respondents seemed to answer "Allow" which is reasonable. The No.Times factor is associated with Japanese respondents' interests. However, the Japanese results might not be applied to Thai respondents.

Table 4.5: Percentage of Japanese and Thai respondents who answer "NOT Allow" when considering the n_0 and n_1 features

n_0	n_1					
	Ads_Part-time job		Ads_Discount price		Game	
	Japan	Thai	Japan	Thai	Japan	Thai
Meeting_5-15 co-workers	**83.33%**	**94.77%**	**91.84%**	72.75%	79.31%	78.22%
Work_Program Analysis	53.33%	**90%**	**85.71%**	50%	80.01%	**81.48%**
Work_Code programming	66.67%	**82.61%**	**88.89%**	45.45%	75.12%	**80%**
Work requirement	**90.91%**	**92.31%**	**85.19%**	76.33%	20.04%	**80.65%**
Private time	62.41%	**86.88%**	62.41%	51.8%	45.77%	64.21%
Travel	**86.67%**	**95.24%**	**95.22%**	59.04%	81.25%	37.5%

On the other hand, Thai respondents tended not to allow, especially when seeing the information 10 times. This indicates that they did not want the information to be fed into their SNP, although they were interested in that particular kind of information. This is because they were concerned about privacy and did not want to reveal their preference. Viewing this kind of post excessively on the SNP gave rise to boredom.

3. **Influence of the age factor on Japanese and Thai respondents' decision**

   The age of the respondents can be divided into four groups for Japanese and three groups for Thais (There are no Thai respondents younger than 20 years old). The results exposed that with increasing age, Japanese and Thai respondents predominantly answered: "Not Allow", especially Thai respondents older than 30 years old as demonstrated in Figure 4.8.

Nonetheless, an attempt was made to find the reasoning behind the respondents' decision by using the n_0 and n_1 features. The respondents older than 25 years old were selected for analysis because the number answering: "NOT Allow" was over 50%. In Table 4.6, the information category (n_1) contained the advertisement and

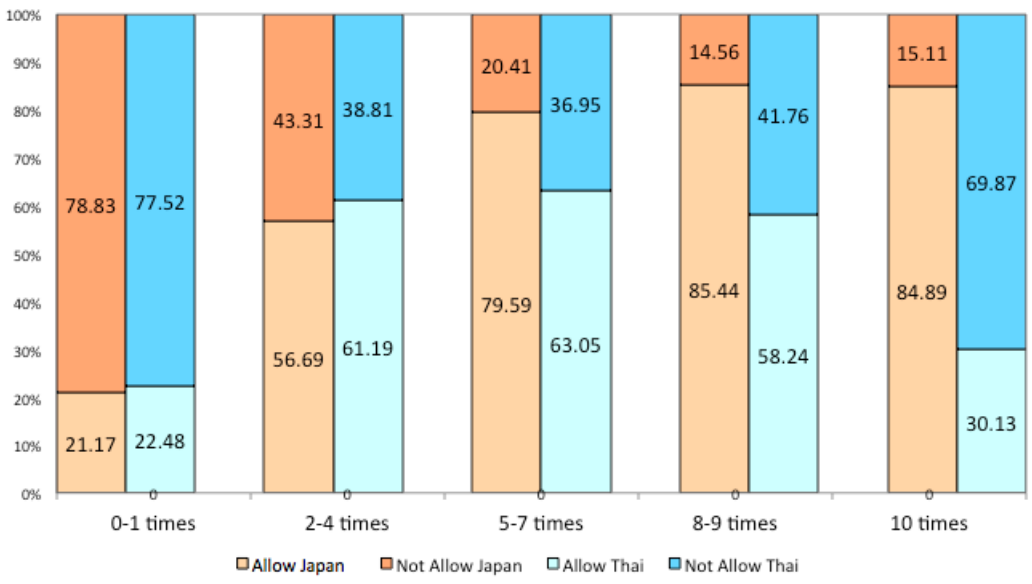


Figure 4.7: A comparison of the No.Times factor on Japanese and Thai respondents' decision

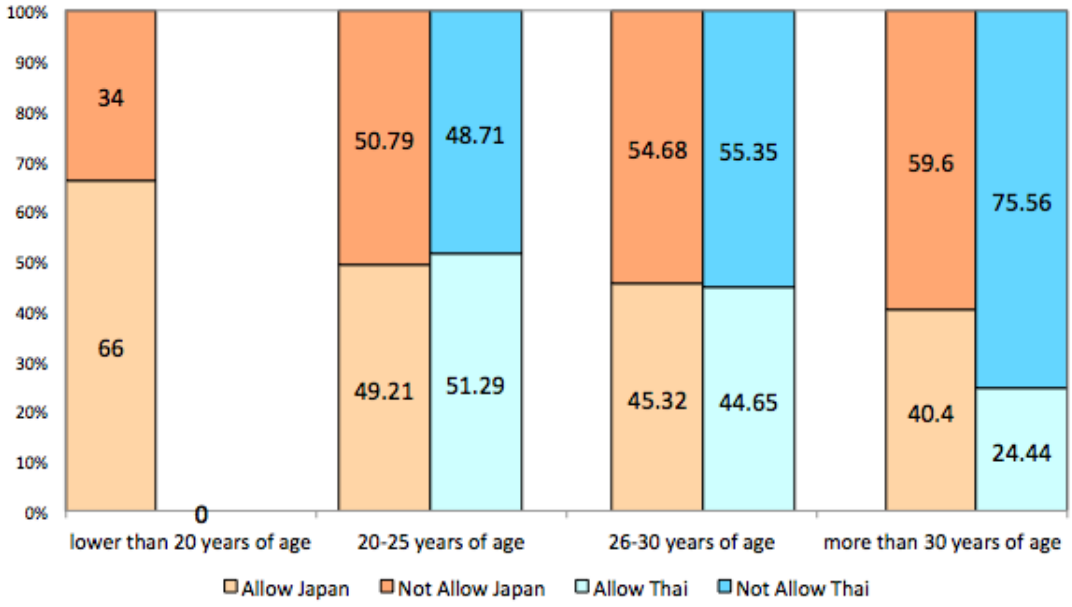


Figure 4.8: A comparison of the age factor on Japanese and Thai respondents' decision (there are no Thai respondents younger than 20 years old)

game, creating problems for the respondents, while other information categories did not influence their decision.

The results clearly showed that on average 86% of Thai respondents were not interested in the advertisement (for a part-time job) at any time. Also, it should not have appeared on the SNP while Japanese respondents had a meeting with 5-15 co-workers and enjoying activities (e.g. shopping, partying, traveling). Moreover, the advertisement involving discount prices was viewed as spam when over 90% Japanese respondents saw this kind of information during meetings and at work. Accordingly, OSNs should emphasize that feeding advertisements and games into the SNP might cause annoyance to readers older than 25 years of age, when their current situation is different.

Furthermore, respondents from both countries were interviewed. From the interviews, it was established that there is a relationship between age and preference. For Thai respondents, the young people interviewed had several interests and were open to reading different things on the SNP. However, the older people had specific interests and pointed to them during their participation in OSNs. For the Japanese, most of them had their own specific interests. One 60 year old Japanese respondent said she was new to OSNs. She tried to use them for several purposes, such as finding new friends, and contacting others without using her cell phone. However, she did not like her SNP to contain large amounts of information as it took her too much time to read and skip that which did not relate to her interests.

4. **An Influence of the time decay (n_6) feature on Japanese and Thai respondents' decision**

Table 4.6: Percentage of respondents older than 25 years of age giving the answer: "NOT Allow", when considering the n_0 and n_1 features

n_0	n_1							
	Ads_Online business		Ads_Part-time job		Ads_Discount price		Game invitation	
	Japan	Thai	Japan	Thai	Japan	Thai	Japan	Thai
Meeting_5-15 co-workers	75.75%	87.96%	**89.19%**	**93.22%**	**90.48%**	74.93%	87.5%	76.1%
Work_Program Analysis	80%	55.87%	69.57%	**86.55%**	**94.59%**	59.26%	76%	80.51%
Private_One's self	21.43%	69.67%	56%	**84.3%**	62.5%	54.86%	44.44%	66.82%
Other(shopping, party and during travel	42.31%	52.42%	**92.31%**	**82.43%**	75%	56.1%	65.22%	39.72%

For the n_6 feature, most of the Japanese respondents allowed the information shown as just having been created on the SNP because they liked to discover the latest information as quickly as possible according to the time series. In Japanese culture, time is important from social aspects, i.e. transportation, business, etc. For instance, railway service is very punctual [100]. In other words, the train always comes to the station on time. Even though the Shikansen does not travel at world record speed, passengers can tell the time by the Shinkansen's arrival at the station. On the other hand, Thai culture views time flexibility as being acceptable in certain situations [109].

5. **Influence of the career factor on Japanese and Thai respondents' decision**

   The n_0 and n_2 features were found to be important to careers when selecting the information to read. In both cultures, engineers and IT respondents were respected at work. They restricted receipt of posts during meetings and at work, especially from family [110]. Japanese culture in the workplace is taken more seriously than in Thai culture. It corresponds to Hofstede's individualism dimension [41], where people living in the same society focus on the interests of individuals rather than groups. It explains that skill and performance are criteria for task assignments in Japan. Promotion relies on the seniority rule, which recognizes age [111]. The Japanese have a famous loyalty to the company, so they seldom change jobs. For Thais, a task is generally assigned to a group. Position and promotion depends on performance, and work periods are important. This leads to competition for new positions or careers.

6. **Further analysis**

   Certain factors were found that may impact on the respondents' decision, but not as much in comparison to the above analysis. For matching between the respondent's preference and the n_1 feature, most Japanese and Thais showed the same cultural behavior if the information categories were the same as the respondent's preference, and around 60-70% of respondents allowed that information to show on their SNP. Moreover, some of them stated that sorting the information on the page was important. Some respondents often have difficulty in finding interesting information to consume because they usually read the information that matches their interests.

Gender did not have much effect on the Japanese and Thai respondents' decision.

4.6 Set of influential features and factors

The influential features and factors obtained from the data analysis component in Section 4.5 are important and necessary for the information filtering component as depicted in Chapter 5. Thus, this section gives an explanation of how each influential feature and factor can benefit the reader when applied to real-situations. In Table 4.4, Section 4.5.1, there are four different features and four different factors for Japanese and Thai.

- **Influential features**

  1. **Reader's current situation**

     Using this feature makes the reader consume information based on a particular situation. Different situations influence the information category that the reader wants to consume. For example, information about games should not be fed during meetings or work time. Additionally, when the reader is traveling for relaxation purposes, information about work should not be displayed on the reader's SNP. This is because such information might change the reader's attitude from being relaxed into being stressed.

  2. **Information category ($\mathbf{n}_1$)**

     This feature is relevant to the reader's preference. If the IFM establishes the information category, it can reduce the amount of information consumed by the reader, thus saving time. For instance, the reader likes football. He/She tends to read sport related information. Hence, information concerning music should not be shown.

  3. **Type or groups of relationship between the reader and creator ($\mathbf{n}_2$)**

     This feature helps the reader to consume information based on the priority of the creator. For example, Reader A gives his/her boss the highest priority so when information is posted by the boss, it should be at the top of the page in order to be easily recognized by the reader. It is possible that Reader B, who has not stayed with his/her family for a long time, is interested in information posted by a family member.

4. **Time decay (n_6)**

   Time decay indicates how recent the information is. Thus, it helps the reader consume the latest information. For example, in Twitter, information is sorted by chronological order. It keeps the reader up to date. Several years ago, one Thai superstar managed to escape from a fire by using Twitter. This is because she immediately asked for help on Twitter. Her tweet was shown at the top of the page and seen by many followers.

- **Influential factors**

  Most influential factors arise from the reader's profile, namely age, career, and preference. Information shown on the SNP will rely on the personality of the reader. For example, the reader's preference can change all the time. The information should be dynamically fed according to preference. The No.Times factor reduces the amount of similar information. For example, there is a political issue in Thailand, thus many creators in OSNs share information regarding political opinions. However, some readers are bored and do not want to see this kind of information.

4.7 Discussion

The analysis results show differences and similarities between Japanese and Thai cultures. The influential features (n_0 and n_1), and the influential factors (No.Times and age) have a high influence on the Japanese and Thai respondents' decision respectively. However, when analyzing in detail, it was found that most Japanese and Thai respondents have different thoughts and feelings. This is because these two countries do not have much shared history. For example, a similar law can provide different results depending on the country. It is a fact that the way people think, feel, or act in a particular country cannot be changed [41].

For example, most Japanese and Thai respondents give "meeting" and "working" as important. This corresponds to the UAI dimension defined by Hofstede [41]. When comparing the UAI score within Asian countries, i.e. Singapore, Hong Kong, etc. Japan and Thailand are in a high UAI score group with 81 and 64 respectively. A country with a

high UAI score in the work place tends to have strict rules and regulations for workers to encourage them to not work hard, considering that time is money. Therefore, feeding information based on the current situation allows the reader to receive suitable information at appropriate times.

However, there are some differences between both cultures. A job in Japan is much stricter than in Thailand and most Japanese respect company discipline. The working style in Japan means that all hierarchy levels have to be involved in every decision. Foreigners view organizations in Japan to be slow in decision making. On the other hand, the working style in Thailand is a top to bottom style, which means that decisions are mostly taken at an executive level (i.e. CEO, project manager, etc.). Tasks are then assigned to their subordinates. Therefore, it is necessary to have a cultural understanding of the work place. However, the n_1 feature is also important, especially with Thai respondents. This feature is related to the respondents' preference. Thus, if information is fed by recognizing the n_1 feature and preference, the post will be more attractive to read. The No.Times factor is an indicator that Japanese and Thai respondents have different attitudes. If the No.Times factor increases, most Thai respondents state that they do not want others to know their preference by seeing the information on their SNP. Excessively seeing this kind of information on the SNP also makes them bored. The age factor shows differences between Japanese and Thai in that when age increases, the use of OSNs might change according to preference and purpose.

The analysis results can be applied to a marketing plan, recommendation, and other cultures

- Marketing plan

  The companies can make good strategic decisions when they want to promote their products to thousands of people by analyzing the best suitable time for distributing their product advertisement in order to be recognized by as many people as possible. Companies can achieve their goals without cost.

- Recommendation

The analysis results can be used to find specific groups of customers so as to recommend products for that group. For instance, companies know that a group of readers enjoy playing football, and they therefore suggest football equipment to this group of readers.

- Applying to other cultures

  The cultural study in this research can be adopted for other cultures or societies with similar characteristics to Thailand and Japan, such as lifestyle, business negotiation, or working practices. People in Southeast Asia share cultural traits, social freedom, and climate. For example, Thailand and Vietnam are similar in the area of business negotiation.

4.8 Conclusion

This chapter investigates the cultural differences between Japanese and Thais focusing on information consumption in OSNs by using a survey. Even though both countries are located in the continent of Asia, Japanese and Thais do not have many shared ways of thinking, feeling, and behavior. The questionnaire in this chapter describes many kinds of scenarios based on a set of pre-defined features related to the cultural differences in OSNs. From the results in Section 4.5.2, they show how the features and factors influence the respondents' decision, providing details from the survey and the interviews. Furthermore, the analysis results can be applied to several aspects, such as marking plans, recommendations, and other societies.

Chapter 5

Culture-based preferences for the Information Feeding Mechanism

5.1 Introduction

Following on from Chapter 4, the set of features and factors that influence the readers' decision making as to whether or not they allow certain kinds of information to be fed into their Social Network Page (SNP) is shown. This is used to make a new repository as a training set. The advantage of this training set is that it can be used to fit a classification model for the information filtering component in this chapter.

The objective of this chapter is to develop a new type of Information Feeding Mechanism (IFM) to reduce information overload with consideration of culture dependency. The function of the proposed IFM is to serve interesting information to the reader dynamically and sufficiently. Therefore, the amount and quality of information depends on it. The reader encounters excessive information and this information is not consistent with their culture. This is because existing IFMs use the same algorithm for every reader to serve information to the reader's SNP, although the readers may come from different cultures. Differing culture preferences can cause information overload [20], and hence it is necessary to construct an IFM for the reader based on his/her culture because there is no universal IFM for all cultures. Each culture should be treated individually.

5.2 Architecture and methodology

Figure 5.1 demonstrates architecture for the core of the proposed IFM that consists of information filtering, repository of training set, data aggregation, OSNs, and information organization. The main component is information filtering. It filters uninteresting and unimportant information by using the classification algorithm. In this research, three classification algorithms: Decision Tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB), are compared to the performance by using classification accuracy and time complexity as the measurement. The inputs of the information filtering component come from two components: repository of training set and data aggregation.

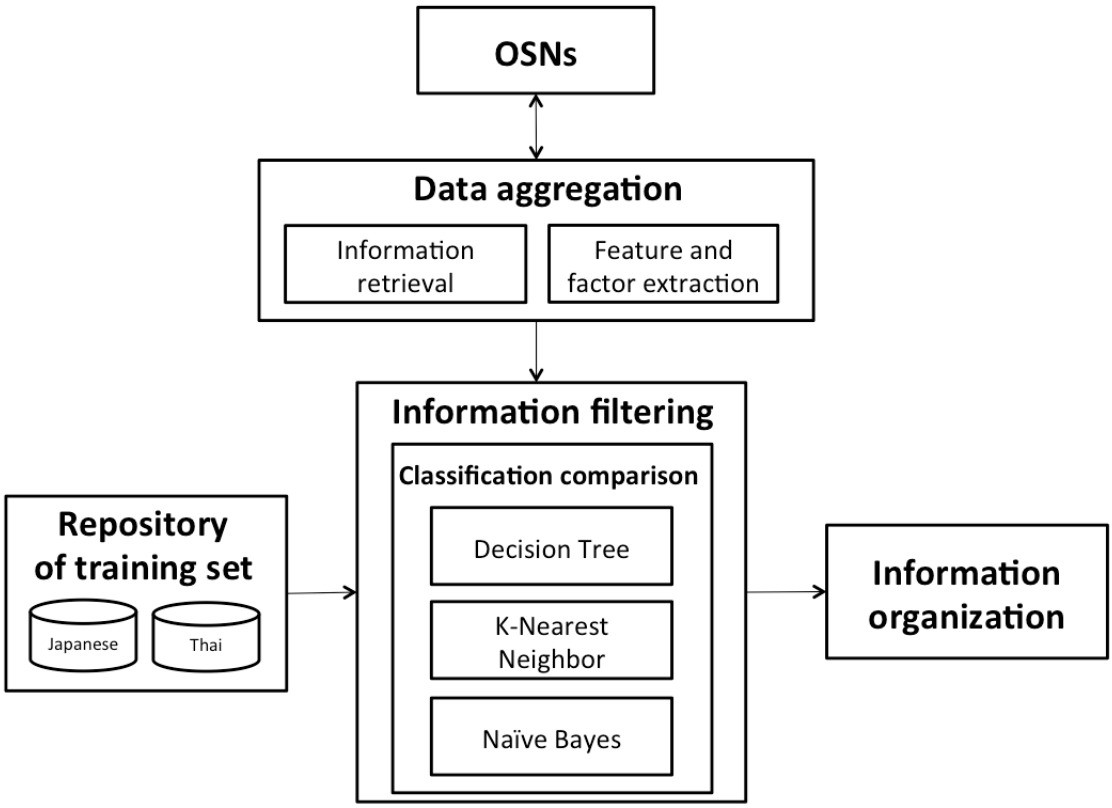


Figure 5.1: The proposed IFM for information consumption in OSNs

Repository of training set is obtained from adjusting the data collected, and the set of influential features and factors as explained in Chapter 4. It is used as the training set in the information filtering component. *Data aggregation* has a duty to find each feature and factor, such as the reader's current situation (n_0), information category (n_1), age, and so on. *Information organization* represents the order of information by using

the reader's preference. This differs from existing IFMs, which orders the information by reverse chronological order or top stories.

5.3 The proposed Information Feeding Mechanism

This section describes how each component in the proposed IFM works and connects so as to filter uninteresting information and serve interesting information to the readers dynamically and sufficiently.

5.3.1 Repository of training set

This consists of the two databases for Japanese and Thais respectively. It is built for use in the information filtering component. Fields in the database of each country are different depending on the set of influential features and factors as illustrated in Table 4.4 in Chapter 4. Therefore, each database will comprise six fields. In addition, each database will be used to find a suitable classification algorithm, obtained from comparing three classification algorithms: DT, KNN, and NB.

5.3.2 Data aggregation

This component is used to prepare a set of inputs including a test set for the information filtering component. The set of inputs is based on the set of influential features and factors in Table 4.4 in Chapter 4. Hence, this subsection gives an explanation of how to obtain each influential feature and factor including the test set. This component consists of two sub-components: information retrieval, and feature and factor extraction.

1. **Information retrieval**

   This research uses real data from the reader as a *test set* in the experiment via Graph Application Programming Interface (API) in Facebook developer [112]. Facebook Query Language (FQL) is used for retrieving the data. The FQL allows the Facebook user to query the data via a SQL-style interface [113]. The reason for using Facebook API is that Facebook is currently the most popular OSN used by many people in different parts of the world. Meanwhile, the tweet in Twitter is limited by

the number of characters (with a maximum length of 140 characters) and LinkedIn has been designed for professional occupations. Hence, they cannot present the real expression of the creator in daily life. However, permission from the reader must be granted as depicted in Figure 5.2 before reading their information in Facebook due to privacy concerns. The data retrieved is kept in separate databases for Japanese and Thais.

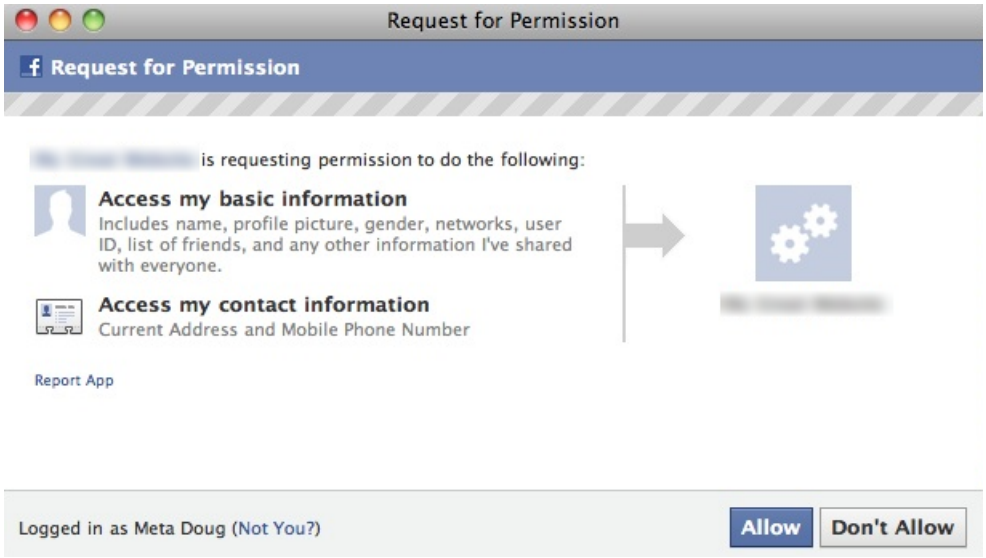


Figure 5.2: Request for permission before reading the information

2. **Feature and factor extraction**

After retrieving the readers' information via Graph API in Facebook developer, there are a lot of properties in this information which are irrelevant for filtering uninteresting information in the proposed IFM. For example, information ID, icon, and place (location associated with the information) are not used for filtering. Therefore, only the necessary information properties have to be extracted in order to use the information filtering component based on the set of influential features and factors from Table 4.4 in Chapter 4. There are four different features and four different factors for Japanese and Thais as presented in Table 5.1. The data relating to each feature and factor can be obtained from three sources: reader, retrieved information, and social graph.

- **Reader source**

Table 5.1: Sources of influential features and factors

Feature/Factor	Influential feature	Source
Feature	Reader's current situation (n_0)	Reader
Factor	Preference	
Factor	Age	
Factor	Career	
Factor	No.Times	
Feature	Information category (n_1)	Retrieved information
Feature	Time decay (n_6)	
Feature	Type or group of relationship (n_2) between the reader and the creator	Social graph

The data relating to features and factors can be queried by using FQL. Before information retrieval, knowledge of FQL commands is required. Data returned from an FQL query is represented in JSON format by default, as demonstrated in Figure 5.3. The results obtained from using the reader source in the information filtering component will rely on the personality of the reader. Furthermore, it is an easy way to filter uninteresting information because it does not need other user ratings to find the similarity between the users and provide suggestions.

- **Retrieved information source**

  The time decay (referring to created time) can be found by using FQL, which is the same process for retrieving the features and factors data from the reader source. However, the information category cannot be obtained from the retrieved information by using the FQL.

  In order to find the information category, it requires a text classification technique. Text classification is used to assign one or more labels (topics, classes) from a pre-defined set into a document or message [114]. It is used in many applications, such as spam filtering, identification of document genre, and so

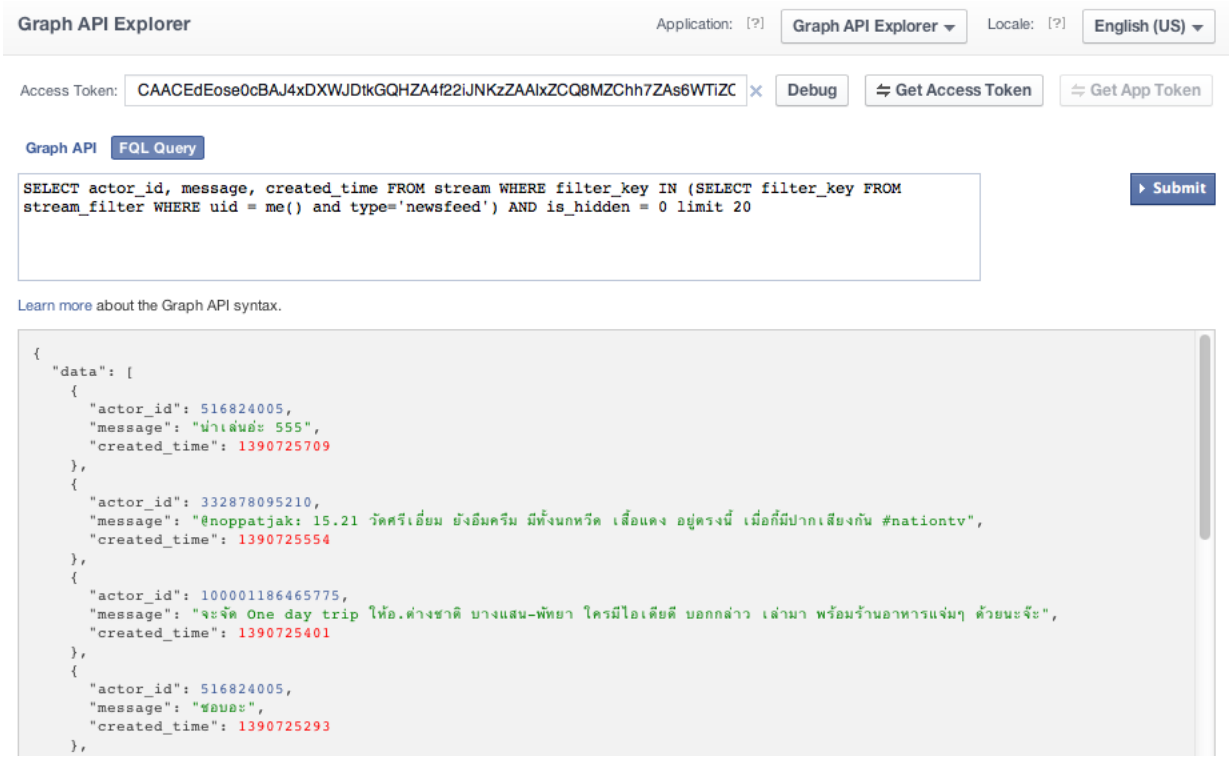


Figure 5.3: Example of a FQL result

on. However, text classification carried out by humans is not only expensive, but also takes a long time to finish. Therefore, assigning the label into the document or message from the pre-defined set is not an easy task. Generally, supervised learning is used for classification because it provides information on the correct classification of documents [115], such as DT, Naïve Bayesian classification, and Support Vector Machines (SVM). This research focuses on category classification based on the information in OSNs as shown in Figure 5.4. This part is conducted in the Thai language only. The information category written in the Japanese language is classified by humans (Japanese). Figure 5.5 shows the steps of category classification, and mainly comprised of six steps as explained below.

A set of category keywords is a major term that represents the category of information. For instance, a music category keyword might be an artist's name, a song, recording (music company) and so on. Keyword preparation is an important task because the corpus construction relies on these keywords.

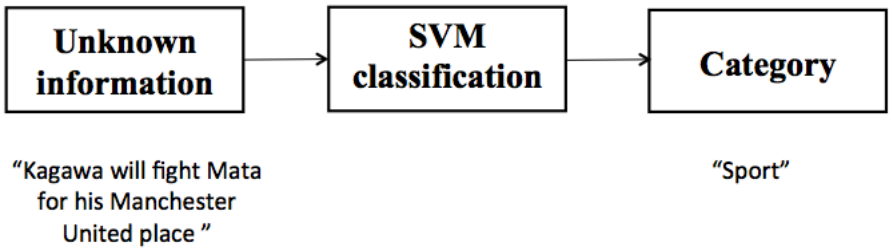


Figure 5.4: A high-level text classification

Corpus construction is a process of making a database for a training set and test set and is used in classification. The data in corpus is obtained from searching by reference to the set of category keywords.

Data cleaning is used to fill-in missing values, identify outliers, smooth out noisy data, and correct inconsistent data. Some retrieved data has to be removed because it is not consistent with the desired category, and it might reduce classification performance. The symbols, such as '@', 'RT' and '#' must also be deleted.

Word segmentation stands for transforming a sentence in the corpus into a sequence of words. This process is used because the Thai language is written continuously. There is no explicit marker between words in a sentence, such as white space, commas as in English, Spanish, German, etc. The longest matching technique based dictionary is applied because it is less complex and powerful.

The preparation of features, training set, and test set are used to format data in the corpus for the SVM classification. It starts from removing the function words, which have little meaning or an ambiguous meaning (e.g. prepositions, pronouns, auxiliary verbs). Then, a list of words or a unigram is considered as the feature set (the unigram obtains the highest accuracy from an experiment). Thereafter, the data in the corpus is set as SVM syntax.

SVM Classification is supervised learning used for data classification. It aims to predict a correct result in the test set in a model produced from the training set. The SVM model represents data in the training set as points in space, and then maps that data into the space with clear gaps between categories. Therefore, when a new input arrives, it will be mapped in the space and predict the most suitable category.

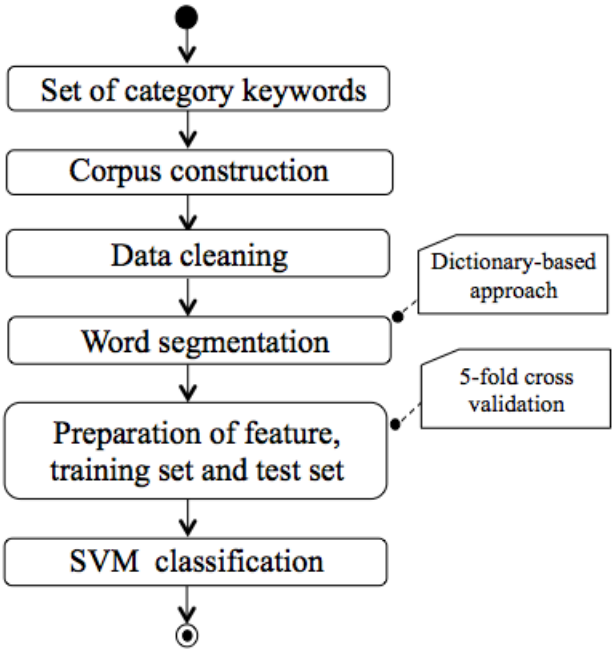


Figure 5.5: Work-flow for category classification

- **Social graph source**

  The type or group of relationship acquired by using FQL. In Facebook, the relationship uses the term "Group" to indicate the reader is a member. Such groups could be family, a group of university friends, co-workers, and so on. The advantage of using a social graph is that it utilizes the data (type or group of relationship) already existing in OSNs, such as "Group" in Facebook and Twitter, "Circle" in Google+, and "Connection" in LinkedIn. The effectiveness of information filtering relies on the successful group management of the reader. Generally, the reader will believe or consider the information posted based on who he/she knows or trusts. For example, if information is posted by the

creator from a group of acquaintances, the reader might ignore it.

5.3.3 Information filtering

This component is considered to be the core of the proposed IFM because it is important for predicting which information should appear on the reader's SNP. Information filtering in this research uses the classification concept to predict a result in the test set using a model produced from the training set. In the classification, many algorithms are powerful and less time consuming. Nonetheless, if the training set is not suitable for those algorithms, the results are not reliable. Therefore, a suitable classification algorithm was found for producing the model by comparing three algorithms: DT, KNN, and NB. Reasons for selecting these three algorithms as candidates are indicated in Table 5.2. The criteria used for algorithm selection are classification accuracy and time complexity. The performance of the selected algorithm is measured against existing research works.

1. **Algorithm selection**

   - **Classification accuracy**

     The classification tool, known as WEKA, is used [105] to measure the performance of three algorithms. For classification, it uses the training set based on different sets of features and factors for Japanese and Thais in Chapter 4, Table 4.4. Each feature and factor are added into three algorithms to observe the accuracy improvement of combinations when the number of features and factors changes. A test option has been set for 10-fold cross validation: the technique used to separate the data into the training set and test set. This technique can avoid problems with over-fitting.

     There are three combination groups: combinations of pure features and one factor; combinations of pure factors and one feature; and combinations of features and factors. For the combinations of pure features and one factor, the results in Table 5.3 show that pure features (Baseline 1-3) cannot give a high performance, showing accuracy levels of 61.74% and 67.41% on average for Japanese

Table 5.2: The algorithm comparison of Decision Tree, K-Nearest Neighbor, and Naïve Bayes

Classification algorithm	Description	Advantage
Decision Tree	This is a tree structure, and is the combination of mathematical and computational techniques.	This has a high predictive performance and can handle a variety of input data: nominal, numeric and textual.
K-Nearest Neighbor	This is a simple predictive algorithm using an entire training database as the model and decides classification using Euclid distance.	This can predict discrete attributes and continuous attributes. It does not require a model to represent the statistics and distributions of the original training set.
Naïve Bayes	This is based on conditional probabilities by applying Bayes' Theorem. The probability is counted by the frequency of values and combinations of values in the historical data.	This affords fast, highly scalable model building and reliable classification performance.

and Thais, respectively. However, other combinations of pure features and one factor (No. 2-5, No. 7-9, and No. 11-13 in Table 5.3) provide higher classification accuracy. When each factor (No.Times factor, age, preference, and career) are added into each Baseline: 1, 2, and 3, the accuracy improves. Especially, when the No.Times factor is extended into Baselines 1-3, the accuracy clearly increases by approximately 5-10%. This corresponds to the results in Chapter 4, Table 4.4, that this factor has the most influence on reader's decisions. Age, career, and preference factors do not substantially increase.

For *combinations of pure factors and one feature*, Table 5.4indicates that the

performance of combination of pure factors (Baselines 1-3) is higher than combinations of pure features with 75.84% and 71.41% on average for Japanese and Thais. It was found that the number of factors affects the improvement accuracy for Japanese and Thais. Three factors give a better performance than two factors. The feature extension has little influence on the accuracy for Japanese respondents. For Thais, adding one feature into pure factors makes little improvement to the performance. It was observed that increasing the n_1 feature into the No.Times, age, and career factors gives a better result than adding the n_0 or n_2 features. This is because the n_1 feature has the highest impact on audience decision as shown in Chapter 4, Table 4.4 . In conclusion, the number of factors affects improvement accuracy and three factors can enhance performance. On the other hand, the feature extension does not have much influence on classification accuracy.

In the *combinations of features and factors*, it was observed that they can obtain higher classification scores than the two previous combinations. As shown in Table 5.5, increase in accuracy comes from the number of combined features and factors. For example, when the preference factor is added into combination Nos. 1 and 4 for Japanese respondents, the classification accuracy is higher.

Based on the results of three combination groups in Tables 5.3-5.5, the NB algorithm obtains the highest classification accuracy with 72.32% and 71.11% and a standard deviation of 5.02 and 2.13 on average for Japanese and Thais respectively, while the KNN algorithm has the lowest classification accuracy as shown in Table 5.6. Nonetheless, the average classification accuracy of three algorithms is not very different and hence computational complexity is used as an evaluation indicator.

Table 5.3: Accuracy of classification for three algorithms when considering combinations of the features and one factor

Combination of the features and one factor	**NB**		**DT**		**KNN**	
	JP	**TH**	**JP**	**TH**	**JP**	**TH**
1. n_0/n_1 (Baseline 1)	64.04%	67.74%	61.21%	65.52%	60.64%	65.52%
2. n_0/n_1+No.Times	73.92%	70.38%	73.97%	71.45%	72.27%	70.36%
3. n_0/n_1+Age	64.48%	69.05%	61.43%	66.53%	59.90%	65.66%
4. n_0/n_1+Preference	67.70%	-	64.91%	-	64.04%	-
5. n_0/n_1+Career	-	67.95%	-	67.20%	-	69.18%
6. $n_0/n_1/n_6$ (Baseline 2)	63.65%	-	60.60%	-	60.30%	-
7. $n_0/n_1/n_6$+No.Times	73.62%	-	73.70%	-	72.18%	-
8. $n_0/n_1/n_6$+Age	64.65%	-	61.21%	-	57.20%	-
9. $n_0/n_1/n_6$+Preference	67.44%	-	64.08%	-	64.20%	-
10. $n_0/n_1/n_2$ (Baseline 3)	-	68.76%	-	68.78%	-	68.15%
11. $n_0/n_1/n_2$+No.Times	-	70.18%	-	71.19%	-	71.35%
12. $n_0/n_1/n_2$+Age	-	69.40%	-	65.98%	-	65.90%
13. $n_0/n_1/n_2$+Career	-		-	66.48%	-	66.85%

- **Time complexity**

  In this section, the time complexity of three algorithms is evaluated to build a model and test the model as denoted in Table 5.7, where m represents the features and factors, n is the number of instances and k is number of nearest neighbors. The time complexity is commonly expressed by Big O notation and shows the time required by an algorithm in the worst case scenario for a given amount of instances. The time taken for each algorithm to build and test models from the classification tool is investigated and shown in Figures 5.6 and 5.7. The number of instances in the training set and test set were simulated by using 10-fold cross-validation. The data is separated into 10 pieces. 9 out of 10 pieces are assigned as a training set and the remaining piece is used for a test set. Then, when the number of instances in the training set and test set increases, the time taken is investigated. The time complexity of the NB

Table 5.4: Accuracy of classification for three algorithms when considering combinations of the factors and one feature

Combination of the factors and one feature	**NB**		**DT**		**KNN**	
	JP	**TH**	**JP**	**TH**	**JP**	**TH**
1. No.Times/Age (Baseline 1)	73.70%	70.15%	75.19%	70.16%	75.40%	69.90%
2. No.Times/Age+n_0	73.10%	70.86%	72.05%	73.31%	73.27%	72.57%
3. No.Times/Age+n_1	73.23%	71.85%	72.18%	73.61	72.18%	71.45%
4. No.Times/Age+n_6	73.18%	-	72.75%	-	72.93%	-
5. No.Times/Age+n_2	-	70.93%	-	70.42%	-	69.88%
6. No.Times/Age/Preference (Baseline2)	75.31%	-	76.77%	-	78.66%	-
7. No.Times/Age/Preference+n_0	76.80%	-	77.5%	-	75.71%	-
8. No.Times/Age/Preference+n_1	76.73%	-	75.49%	-	78.41%	-
9. No.Times/Age/Preference+n_6	76.10%	-	77.69%	-	77.02%	-
10. No.Times/Age/Career (Baseline3)	-	72.14%	-	72.51%	-	73.60%
11. No.Times/Age/Career+n_0	-	72.72%	-	72.62%	-	70.91%
12. No.Times/Age/Career+n_1	-	72.86%	-	72.91%	-	72.99%
13. No.Times/Age/Career+n_2	-	72.38%	-	72.72%	-	70.47%

Table 5.5: Accuracy of classification for three algorithms when considering combinations of the factors and one feature

Combination of the features and factors	**NB**		**DT**		**KNN**	
	JP	**TH**	**JP**	**TH**	**JP**	**TH**
1. n_0/n_1/No.Times/Age	75.75%	73.21 %	75.00 %	72.68%	75.7 %	71.01 %
2. n_0/n_1/No.Times/Age/Career	-	74.32%	-	73.31%	-	74.3%
3. n_0/n_1/No.Times/Age/Preference	76.45%	-	80.02%	-	79.67%	
4. n_0/n_1/n_6/No.Times/Age	77.95%	-	77.58%	-	77.81%	-
5. n_0/n_1/n_6/No.Times/Age/Preference	**78.53%**	-	79.71%	-	79.80%	-
6. n_0/n_1/n_2/No.Times/Age	-	71.92%	-	72.91%	-	71.01%
7. n_0/n_1/n_2/No.Times/Age/Career	-	**75.96%**	-	73.68%	-	73.68%

algorithm is linear, while the DT algorithm requires linearithmic time in the training model and logarithm time in the testing model, and the KNN algorithm consumes linearithmic time only in the testing model.

Table 5.6: Average classification accuracy and standard deviation for three algorithms

Nationality	**NB**	**DT**	**KNN**
JP	72.32%±5.02	71.65%±6.74	71.38%±7.43
TH	71.11%±2.13	70.70%±2.88	70.24±2.68

Table 5.7: Time complexity for training and testing a model

Classification algorithm	**Training model**	**Testing model**
NB [116]	$O(mn)$	$O(m)$
DT [105][117]	$O(mn\ log\ n) + O(n(log\ n)^2)$	$O(log\ n)$
KNN [105]	$O(1)$	$O(kn\ log\ n)$

For the time taken to build the model in Figure 5.6, the graph of each algorithm is shown to be clearly different. It shows that the NB algorithm takes little time to compute probability tables from classes, features, and factors. The DT algorithm requires the longest time because it requires many processes for building a tree structure, such as numeric sort, sub-tree replacement and sub-tree lifting, and converting to a set of rules [105]. The KNN algorithm takes time as constant because it does not work until classification time [105].

For time taken to test the model in Figure 5.7, the NB and DT algorithms do not take much time because they apply the training model to predict the test instances. Whereas the KNN algorithm makes predictions by using the Euclid distance between every instance in the test set and entire training set, and finds k, the nearest neighbor in every instance. Therefore, this process takes the longest time.

Based on classification accuracy and time complexity, the NB algorithm is the most suitable to use for the information filtering component. One more reason to support the NB algorithm is that it can be applied to a real data set, representing exponential growth in OSNs. In 2013, Facebook served 2.5 billion photos per week [21]. When considering other types of information, e.g. text, video, and link, the

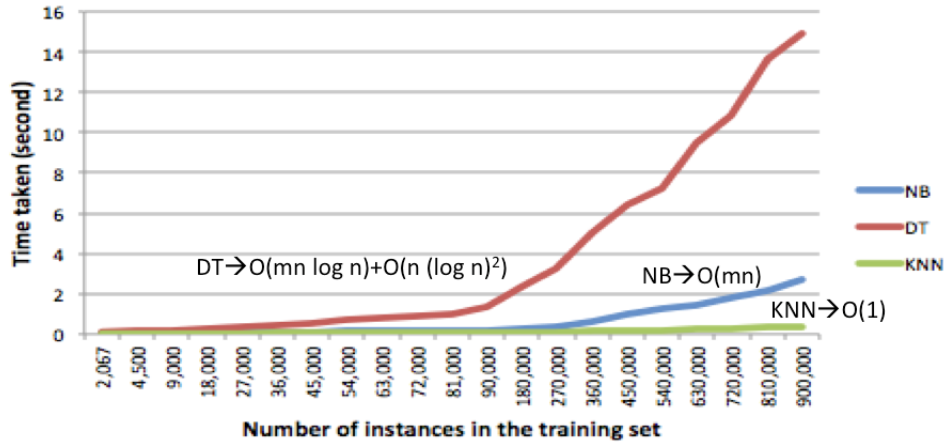


Figure 5.6: Time taken to build model

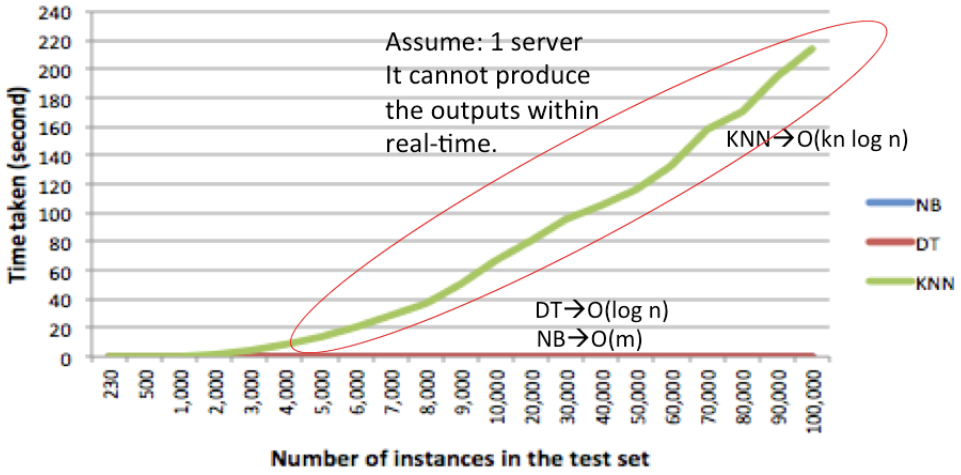


Figure 5.7: Time taken to test model

amount becomes increasingly larger. Thus, Facebook needs parallel computation. The NB algorithm has the possibility to efficiently work on a real data set by using the concept of parallelism [118]. This concept can also be adopted for the DT and KNN algorithms [119][120]. However, when a large amount of data is computed by the DT and KNN algorithms, at some point each server will face the same problem, similar to these graphs in the training model and testing model.

In addition, the reader's current situation feature is important to the proposed IFM, and therefore when it changes, the proposed IFM will compute which information should appear according to new current situations only in that day. This is because

the proposed IFM can re-use sets of previous information never seen before, for the same current situation, and when the reader wants to see more previous information he/she can use the scrolling down facility (re-use concept). For example, yesterday, the reader consumed the information set A during their working period. Today, when the reader's current situation changes from the meeting period to working period, the proposed IFM computes the information set B, posted that day, to show the working period on the SNP. If the reader wants to consume more information during the working period, the proposed IFM provides information set A to the reader. Meanwhile, if the current situation does not change, but new information arrives, the proposed IFM will not re-compute the entire information from past to present. It will calculate only new information because it applies the re-use concept. The type or group of relationship between the reader and creator feature and preference factor are also used in the proposed IFM. When these change, the proposed IFM has to re-compute the entire information. However, it is not a problem when a large amount of information has to be computed because the proposed IFM uses the NB algorithm, which is less complex and efficient. However, it is a problem for the KNN algorithm, when the information shows more than 4,000 instances as in Figure 5.7, and one server might not be able to produce the output within an acceptable time.

In Table 5.5, the set of features (n_0, n_1 and n_6) and the factors (No.Times, age, and preference) are used to achieve the highest accuracy of 78.53% for Japanese respondents. Also, the set of features (n_0, n_1 and n_2) and the factors (No.Times, age, and career) are selected for Thai respondents because it shows the greatest accuracy at 75.96%.

2. **Performance comparison of information filtering**

The concept of performance comparison can be explained by Figure 5.8. This intersecting part is called "virtual existing information filtering". Intersecting features and factors of the proposed information filtering and existing research works are used in the classification. There are four reasons for using intersecting features and factors.

- They are considered as common methods in existing research works.
- They indicate the importance of the proposed features and factors, selected for filtering uninteresting information in this research.
- Some features and factors in existing research works do not relate to cultural differences in OSNs.
- There is some difficulty in setting the same experimental environment and set of features and factors of the proposed information filtering and existing research works.

To measure the classification accuracy of the proposed information filtering and virtual existing information filter, the NB algorithm is used since it is recognized as being the most suitable algorithm for such classification. The summary of performance comparison is denoted in Table 5.8.

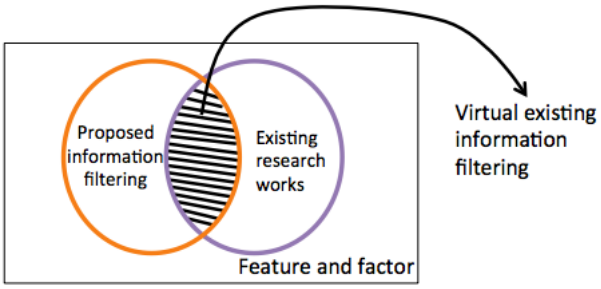


Figure 5.8: A concept for performance comparison of information filtering

Table 5.8: Features and factors for performance comparison of information filtering

Research	Nationality	Algorithm	Feature and factor
Proposed information filtering	JP	NB	Reader's current situation, **information category**, **time decay**, No.Times, **age**, and **preference**
	TH		Reader's current situation, **information category**, **type or group of relationship between reader and creator**, No.Times, **age**, and career
Virtual existing information filtering	JP		Information category [28][97][121], time decay [11][29], age [28], and preference [29][96]
	TH		Information category [28][97][121], type or group of relationship between reader and creator [28], and age [28]

In Figure 5.9, there are significant differences in classification accuracy for the proposed information filtering and virtual existing information filtering using condition $t(58) = 6.535$ for Japanese and $t(47.81) = 3.739$ for Thais at $p < 0.05$ by T-Test. The classification accuracy of the proposed information filtering for both Japan and Thailand exceeds that of virtual existing information filtering. Classification accuracy can be improved by 10% in the reader's current situation feature, the No.Times factor, and the career factor. These features and factors are not commonly used in existing research works, but they are necessary for information filtering. For example, interesting information is dynamically shown on the SNP based on the reader's current situation because the same information might have different importance when the reader is in a different situation. For the No.Times factor, the reader is sometimes bored when consuming a lot of information of a similar content, and therefore this kind of information can be filtered by the No.Times factor. In the career factor, some careers are shown to be restricted by not allowing the use of OSNs in working time, such as in banking, medicine, and so on. Thus, when those in such careers have time to use OSNs, they need to read interesting information.

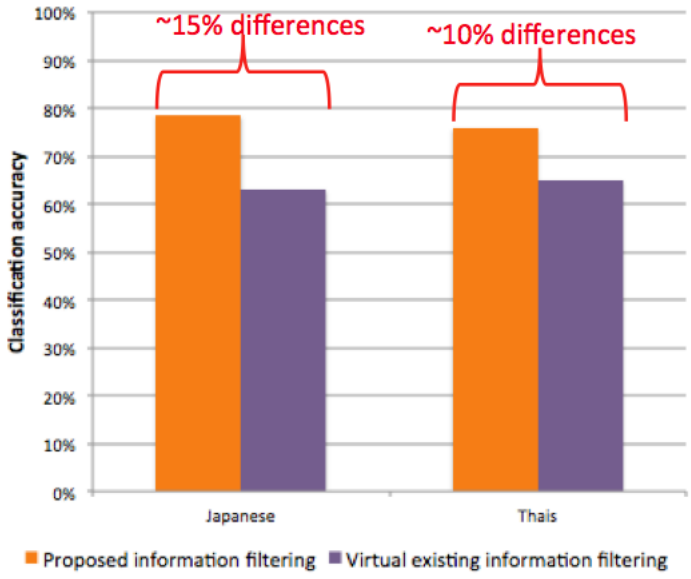


Figure 5.9: Performance comparison of the proposed information filtering and virtual existing information filtering

5.3.4 Information organization

Reverse chronological order and top stories are two of the main information organizational categories in existing OSNs. Reverse chronological order shows the information posted by the creators according to the time it is created, such as Facebook, Twitter, and Google+. The newest information will be presented at the top of the SNP as demonstrated in Figure 5.10. This means the reader always consumes the latest information. Usually, Facebook in using this information organization method displays updated information from 250 friends and Facebook Pages. Top stories in Facebook show the amount of popular information that is of interest to favorite friends. This information organization relies on factors such as the number of comments, as used in Facebook's EdgeRank algorithm [25]. Nonetheless, it is proposed that any information should be dynamically ordered by the reader's preference. This is because the reader's preference can change continually. Moreover, in a real situation, reading an enormous amount of information in the SNP is compared to finding interesting information, and depends on each reader's preference [60]. Therefore, this component allows the reader to set three preferences in descending order. Thereafter, when it orders information, it considers the category extracted in Section 5.3.2 For the information not relating to the reader's preference, this will be sorted in reverse chronological order. Such information organization leads to faster reader consumption of a large quantity of information and is consistent with the requirements of the reader.

5.4 Experimental setup

The goal of the experiment is to evaluate the performance of four IFM methods as shown in Table 5.9 by using questionnaires. The respondents are presented with a virtual SNP to simulate an OSN information consumption situation by showing information selected using four IFM methods.

Four IFM methods use the same test set, which is acquired from Section 5.3.2. Method 1 is a nominal method with no exact specification. The information is selected and shown to the respondent's SNP randomly; hence the respondent cannot expect information characteristics and order. Method 2 applies a timeline and reverse chronological order tech-

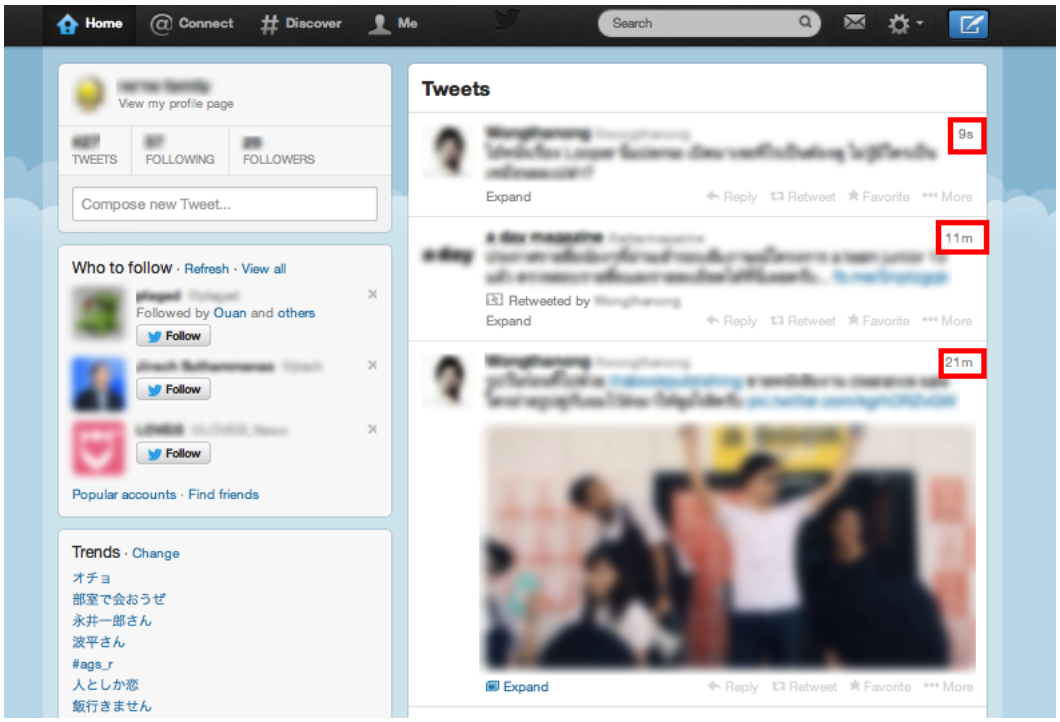


Figure 5.10: An example of reverse chronological order

nique to show the information. Method 3 uses the EdgeRank algorithm for information feeding and top stories for organizing the information. Further details for Method 2 and 3 can be found in Section 2.3.1 and 5.3.4.

Method 4 feeds the information to the respondents by using the NB algorithm together with different sets of features and factors for Japanese and Thais as described in Section 5.3.2. This method uses the training set as mentioned in Section 5.3.1, and the test set for retrieving the respondent's data in Facebook as explained in Section 5.3.2.

The test set is identified to indicate the required features necessary for classification. For example, the information in the test set has to be categorized based on the training set, and is classified by the NB algorithm. Method 4 dynamically serves the information to the respondent, and this information is then ordered by their preference. Furthermore, it allows the respondent to change his/her current situation and preference in relation to their requirements. Figure 5.11 illustrates a prototype of the respondent's SNP after using the proposed IFM based on Bayes' Theorem [105] consisting of three parts: the re-

spondent's profile (name, current situation, and preference), list of information (creator's name, type or group of relationship, information category, content of information, and time created) and results of filtering based on influential features and factors.

To prevent bias, four named methods are not revealed. In this experiment, the real data from each respondent is retrieved via Graph API in Facebook developer [112] as explained in Section 5.3.2. The advantage of using real data is that the respondents do not need to imagine and carry out the experiment due to data familiarity. However, permission from respondents is required before obtaining the data due to privacy concerns.

Table 5.9: Description of information feeding and information organizations

Method	**Information Feeding**	**Post Organization**
1	Random	Random
2	Timeline	Reverse chronological order
3	EdgeRank	Top stories
4	Our proposed IFM	Preference

5.5 Results and Performance evaluation

5.5.1 Questionnaire

After the respondents recognized the differences in four IFM methods, they completed a questionnaire to measure the performance of each method. This questionnaire is answered by 17 Japanese and 22 Thai respondents respectively. 42.4%, 25.7%, 22.9% of the respondents are students, engineers, and IT specialists, respectively. 6.2% of the remaining respondents are other persons, such as teachers, secretaries, and so on. The respondents in the experiment are experienced in the use of Facebook, Twitter, Google+, and Mixi∗ (∗Japanese respondents only). Most of the respondents use OSNs for a specific purpose, such as entertainment, information sharing, consumerism, business, passing the time, and relationship maintenance. In addition, Japanese respondents use OSNs to recruit new

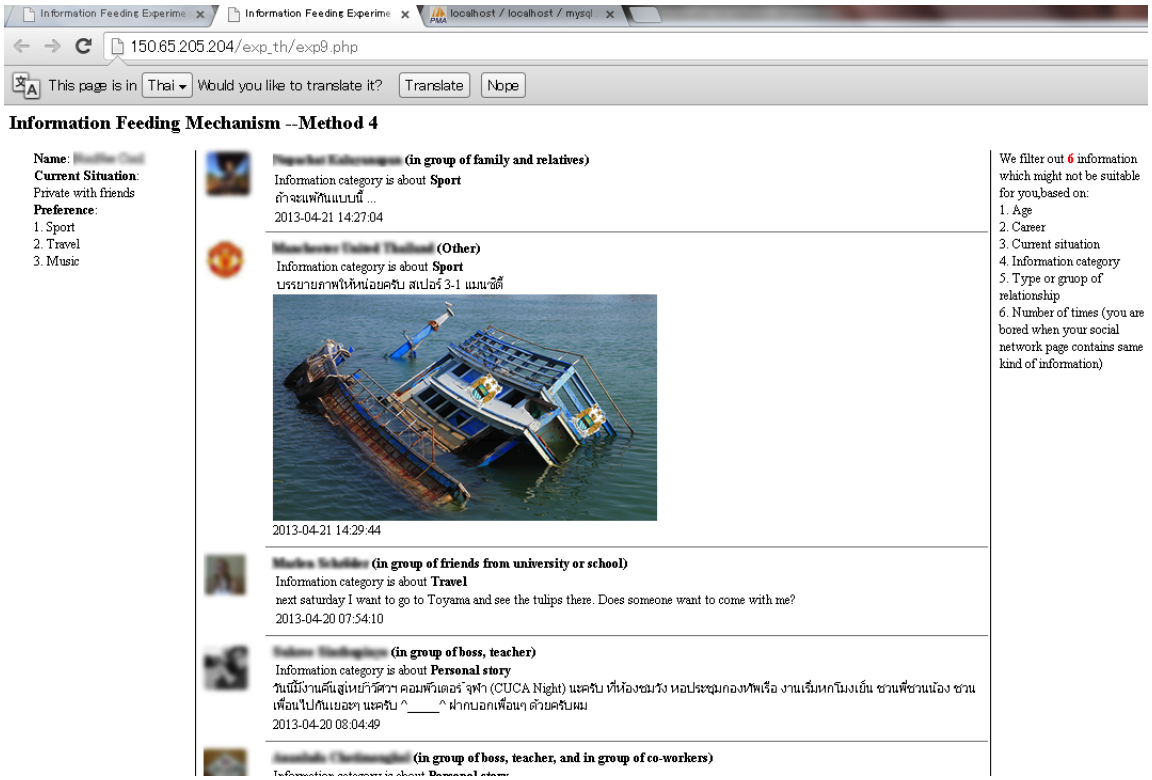


Figure 5.11: An example of the classification results

members. The amount of time spent using OSNs in one day ranges from less than 30 minutes to two hours for Japanese and from one to five hours for Thais. Most of the Thai respondents have more than 200 members in their contact list, while Japanese respondents have 20-100 members. Each respondent is asked 25 questions about demography and use of OSNs, experiences in current OSNs, and the performance of four IFM methods. The questionnaire is shown in Figures 5.16-5.18 at the end of this chapter. However, some questions might be analyzed because they are ambiguous. Each question measures the ability of four IFM methods in Table 5.9 by mean and standard deviation. All questions use Yes/No answers and a 5-point scale (1=0% and 5=75-100%).

5.5.2 Classification results

Figure 5.12 shows the ability of the proposed IFM for filtering uninteresting and unimportant information, when Japanese and Thai respondents have different current situations. The percentage of filtered information in this graph measures the amount of filtered information against the total information and multiplying it by 100. The graph pattern

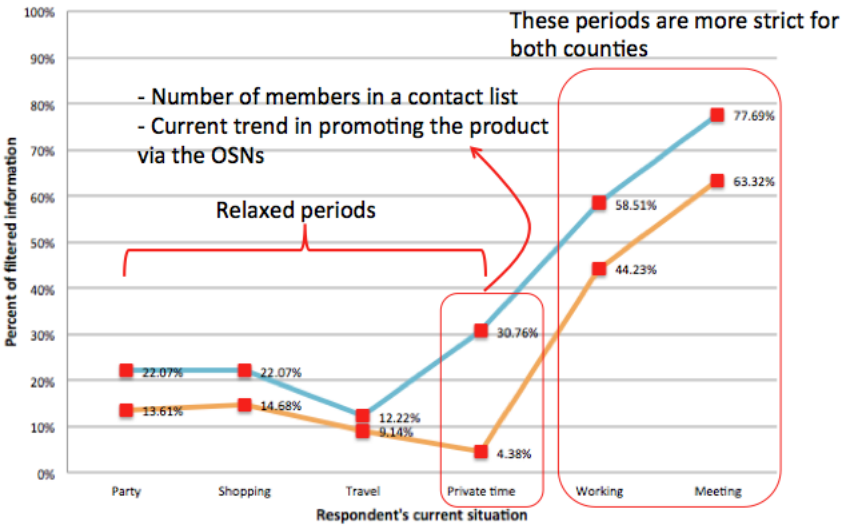


Figure 5.12: Results of classification based on the respondent's current situation

of both countries is quite similar. Firstly, the percentage of filtered information for Thai respondents is higher than that of Japanese in every current situation. Secondly, the percentage of information during meetings (63.32% for Japanese and 77.69% for Thais) and working periods (44.23% for Japan and 58.51% for Thai), is greater than that in relaxed periods, such as partying, shopping, travel, and private time. This corresponds to the culture of both countries in that these periods are quite strict and serious, meaning the respondents concentrate on their tasks. In the relaxed periods, the percentage of filtered information is pretty low, especially regarding the private time of Japanese respondents. This percentage differs considerably from the Thai results. Thai respondents have a higher number of members in a contact list than the Japanese, and Thai creators currently tend to promote their products via OSNs, hence this requires the filtering of uninteresting information. This can be seen by reference to Figure 5.13, which shows that the percentage of filtered information categories between Japan and Thailand, e.g. music, travel, and games does not show a significant gap, except for advertisements. It is clear that 90.2% of advertisements are removed from the Thai respondents' SNP.

Figures 5.14 and 5.15 depict the percentage of information categories filtered in each current situation. This percentage is calculated using the filtered information in each category and current situations against the total amount of information in each category and

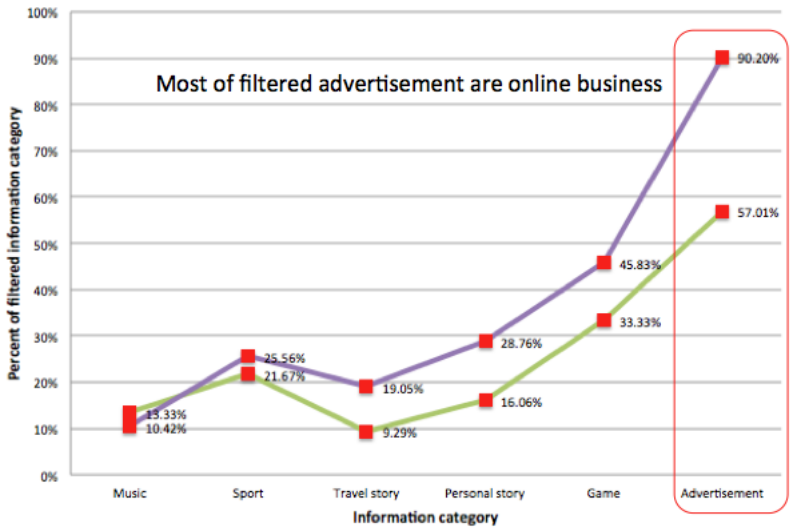


Figure 5.13: Results of classification based on the information category

multiplying it by 100. An overview of two graphs shows the similarities and differences between Japan and Thailand. Similar points apply to meetings and working periods, where information with a high filtering ratio is removed, especially advertisements, games, and sport. This means that when information comes to the respondent during these current situations, most of it will not be shown on the respondent's SNP. This corresponds to the results illustrated in Figure 5.12. The respondents are served by certain specific information, which is interesting and important. No matter what the current situation is, most advertisements are filtered, but the percentage of filtered information in each category for Thailand is higher than that of Japan.

In different aspects, no matter what the Thai respondent's current situation, advertisements are distinctly filtered at a rate higher than 80% on average. From observation, the advertisements collected are about promoting the products for online business. Meanwhile, for Japanese, the advertisement is filtered depending on the respondent's current situation. In particular, Japanese respondents are more likely to read an advertisement in their private time. This differs completely from meetings and working periods, where they do not want to read advertisements. Therefore, controlling what information should appear on the SNP is not an easy task for the proposed IFM when the readers are of different cultures. Also, the results indicate that the reader's current situation and in-

formation category are important for filtering uninteresting information. Controlling the information on the SNP relies on each cultural preference.

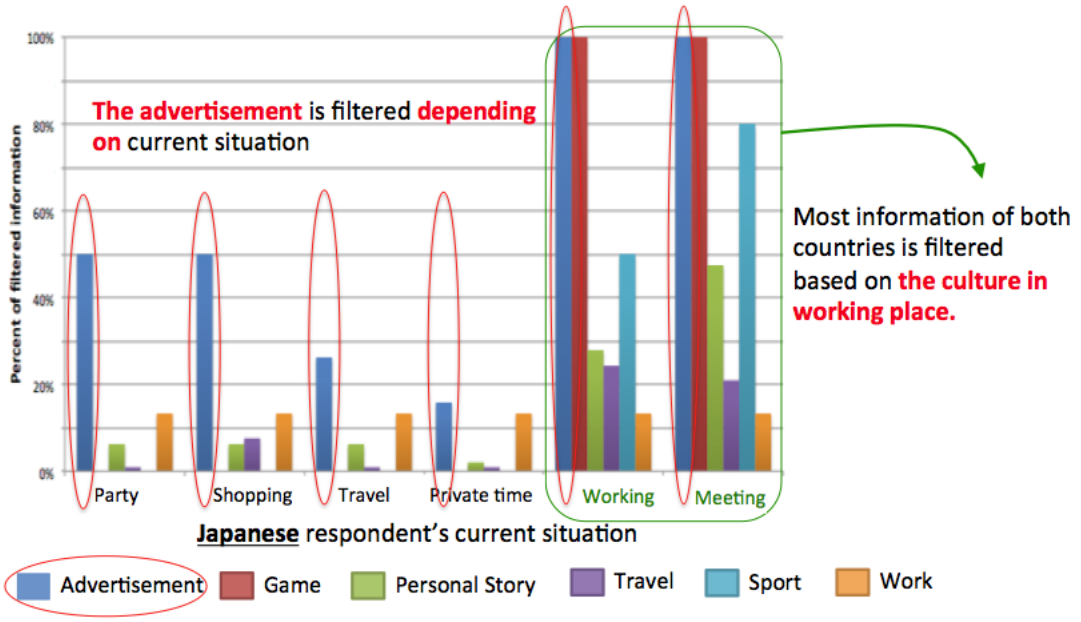


Figure 5.14: Results of classification based on Japanese respondent's current situation and the information category

5.5.3 Opinion of the problems in current OSNs

Respondents are asked about problem experiences in OSNs, such as information overload, missing and consistent information serving. Table 5.10 shows that respondents are encountering information overload on their SNP by information uninteresting to them. When they read too much information, they feel they are forced to read useless information. These results are relevant to Bontcheva et al. [34]. 64.29% and 71.43% of Japanese and Thai respondents express their annoyance at having their privacy disturbed. They miss useful information around 4-6 times a day. The missing information means that it can be blocked by other information, such as nonsense, uninteresting information, etc. Interestingly, 66.67% and 85.71% of Japanese and Thai respondents believe feeding information to the SNP based on their current situation to be useful. However, when multiple choices are provided to the respondents for selecting which IFM can solve the information overload problem, Japanese respondents think that feeding information like Timeline can

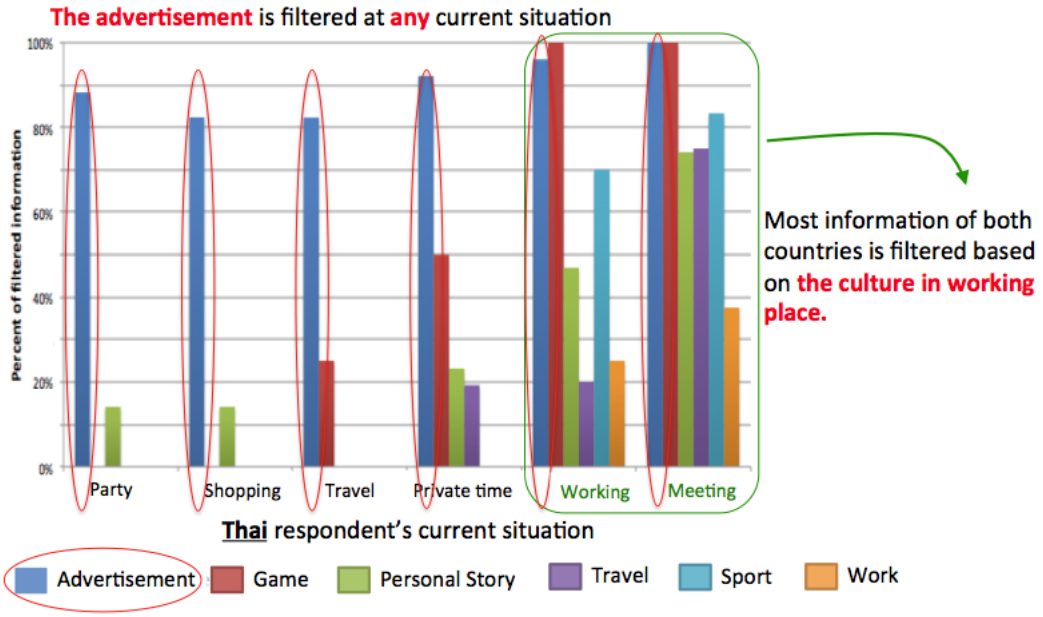


Figure 5.15: Results of classification based on Thai respondent's current situation and the information category

solve the problem at 3.29 ± 0.91, while Thai respondents believe that using current situations can reduce this problem at 3.52 ± 0.98. This indicates most Japanese respondents are more likely to use Timeline as the IFM, whereas Thai respondents prefer information to be selected and displayed dynamically based on their current situation. Information organization also has an impact on the reading content. Some Thai respondents said it orders the information according to priority, with the most significant information shown at the top of the page, whereas uninteresting information is displayed at the bottom of the page.

Table 5.10: The respondents' opinions concerning information overload on the SNP

Variables	**JP**	**TH**
Excessive information	3.36 ± 0.93	3.67 ± 1.11
Unwanted information	3.07 ± 1.00	3.47 ± 1.12
Privacy disturbance	64.29%	71.43%

5.5.4 Performance of four IFM methods

Table 5.11 reveals significant differences between Japanese and Thais. Each question contained by evaluator is asked independently. For overall performances, Method 1 is the worst performance for both countries because it does not use any algorithm for feeding the information. Other methods are described below, which rely on each evaluator because each method has different advantages and depends on the attitude of each culture or life style. Thus, analysis of further details proves interesting.

Table 5.11: Performance evaluation using mean and standard deviation

Evaluator	**Method 1**		**Method 2**		**Method 3**		**Method 4**	
	JP	**TH**	**JP**	**TH**	**JP**	**TH**	**JP**	**TH**
Information overload solution	2.43±1.22	2.43±1.12	**3.29±0.91**	2.62±1.02	2.43±0.85	3.19±0.93	2.64±0.93	**3.52±0.98**
Information filtering performance	2.50±1.22	2.57±1.12	2.79±0.97	2.67±1.11	3.00±1.17	3.10±1.22	**3.43±1.08**	**4.05±0.97**
Dynamical information feeding	1.86±0.86	2.67±1.28	3.36±0.93	2.95±1.12	2.86±1.10	3.14±1.06	**3.43±0.94**	**3.52±1.08**
Consistent information serving	2.93±1.14	2.81±1.33	**3.64±0.93**	3.19±0.98	3.43±0.94	4.04±0.80	3.57±1.02	**4.24±0.89**

- **Information Overload Solution**

  In this evaluation, the ability of each method to solve the information overload problem is considered. For Japanese, after taking the ANOVA test, four methods show no significant differences, $F(3, 52) = 2.352$, $p > 0.05$. Most respondents think Method 2 (3.29 ± 0.91)can solve the information overload problem. This is relevant to the previous analysis in Chapter 4, Sections 4.5 and 5.5.3, where most Japanese give importance to time. The information fed according to time is likely to encourage them to consume such information more quickly and get updated information from their friends and other users. Therefore, most Japanese respondents feel this method can solve the information overload problem. Further analysis shows that age, career, and gender have no influence on the results as depicted in Table 5.12.

  For Thais, there is at least one significant difference among the four methods, $F(3, 80) = 5.21$, $p < 0.05$. Scheffe values show that Method 1 (2.43 ± 1.12, $p = 0.01$) and Method 2 (2.62 ± 1.02, $p = 0.047$) are statistically significantly lower than Method 4 (3.52 ± 0.98). Method 3 is not statistically significant with Method 4 ($p > 0.05$). Most of the respondents think Method 4 is the most appropriate for

solving excessive information feeding in Table 5.5.3 due to its flexibility. The quantity of information can be changed according to the current situation. When age groups and careers are analyzed as shown in Table 5.13, the results reveal that most of the respondents aged between 26 and 30 years old (3.55 ± 0.93) as well as engineers (3.50 ± 1.05), believe in Method 3.

- **Information Filtering Performance** The respondents are asked about the performance of each method in removing the uninteresting and unimportant information according to current situations or needs. The results of all four methods are not statistically significant in this evaluation for Japanese, $F(3, 52) = 1.707$, $p > 0.05$, while four methods show significant differences for Thais, $F(3, 80) = 7.755$, $p < 0.01$. A post-hoc test indicates that Method 4 (4.05 ± 0.97) is significantly higher than Method 1 (2.57 ± 1.12, $p = 0.001$) and Method 2 (2.67 ± 1.11, $p = 0.002$). Method 3 is almost significantly different at $p = 0.06$, when compared to Method 4.

  The overall performance clearly shows that Method 4 (3.43 ± 1.08 Japanese, 4.05 ± 0.97 Thais) overcomes the remaining methods. It filters uninteresting and unimportant information by using the NB algorithm based on the set of influential features and factors. The respondents' current situations were analyzed from the interviews. The results show that most Japanese and Thai respondents need information filtering, especially when at work or during meetings.

  Beside the quality of information, the number of members in a contact list might create the need for information filtering. Presently, in addition to Mixi, most Japanese use OSNs such as Facebook and Twitter, and getting new members is one of the purposes for their use. This leads to an increase in the receipt of useless information. Most Thai respondents have on average more than 200 members in their contact list. Hence, they have a high chance of receiving excessive information and need to filter it.

- **Dynamical Information Feeding**

  This evaluation measures how each method can dynamically serve the informa-

tion to respondents based on their current situation or needs. Four methods for Japanese respondents have statistically significant differences, $F(3, 52) = 7.952$, $p < 0.05$. Method 1 (1.86 ± 0.86) has significant differences to Method 2 (3.36 ± 0.93, $p = 0.002$), and Method 4 (3.43 ± 0.94, $p = 0.001$) has a 0.05 significance level, whereas there are no statistical differences between the results in the four methods for Thais in this evaluation, $F(3, 80) = 2.099$, $p > 0.05$.

However, Method 4 (3.43 ± 0.94 Japan, 3.52 ± 1.08 Thais) shows the highest performance for both countries. The NB algorithm uses the respondent's current situation, which is one set of features and factors of classification, and therefore when their current situation changes, the information fed into the SNP also changes.

Nevertheless, Method 2 for Japanese respondents cannot be discarded because its mean score is closer to Method 4. Most of the Japanese engineers (3.00 ± 1.00), IT specialists (4.00 ± 0.00), and females (3.83 ± 0.41) as presented in Table 5.12 said that when they open OSNs, the information is dynamically fed according to time change. Consequently, they can consume updated information.

- **Consistent Information Serving**

Table 5.10 reveals that the respondent's SNP contains unwanted information, and therefore its quality is important depending on their requirements. For Japanese respondents there are no statistically significant differences among the four methods, F$F(3, 52) = 1.425$, $p > 0.05$. For Thais, the performance of the four methods show significant differences in this evaluation, $F(3, 80) = 9.397$, $p < 0.05$. By running post-hoc tests, Method 4 (4.24 ± 0.89) is statistically significantly higher than Method 1 (2.81 ± 1.33, $p = 0.0001$) and Method 2 (3.19 ± 0.98, $p = 0.015$). Method 3 is also significantly different to Method 1 ($p = 0.003$).

Table 5.11shows that most of the Japanese respondents believe Methods 2-4 are effective in serving interesting and important information by using different concepts. Nevertheless, Method 2 possesses the highest mean score (3.64 ± 0.93).. It was

found that career and age have an impact on the results, as illustrated in Table5.13 Japanese respondents, who are students (3.75 ± 1.04) or aged between 21 and 25 years old (4.00 ± 1.00), think Method 4 has the best performance.

For Thais, although Method 4 (4.24 ± 0.89) still satisfies the respondents, Method 3 (4.04 ± 0.80) cannot be ignored, since it is slightly lower in the mean score than Method 4. Most of the female respondents (4.46 ± 0.52) trust Method 3 as indicated in Table 5.13. The interviews indicate they are usually interested in entertainment or fashion, and therefore the possibility that these are out of date.

From analysis and interview, Japanese and Thai respondents show significant differences in the selection of suitable IFMs for information consumption in OSNs. For Japanese respondents, it is not clear which method is the most appropriate. However, career and age were found to have an influence on the overall performance in four evaluations. Respondents over 26 years old (3.60 ± 0.71), IT specialists (3.75 ± 0.43), or engineers (3.33 ± 1.07) believe that Method 2 helps them to obtain the latest information, which they can follow in real-time. Especially during the working day, when they are very busy, they need to consume the entire information they need by way of a short message, in the fastest possible time. This indicates that time is important to them [100]. Nonetheless, respondents aged between 21 and 25 years old (3.62 ± 1.13), or students (3.67 ± 0.99), like Method 4. They state that they can obtain suitable information about what their friends are doing in current situations automatically and dynamically. They do not want to select any information to read, but they need the IFM to choose it for them. This is because these groups of respondents usually have many different activity periods during the day, such as a class, or seminar, and so on.

Meanwhile, Method 4 clearly satisfies Thai respondents. It can solve the information overload problem because it reduces the quantity of information on the SNP according to the set of influential features and factors, such as respondent's current situation, and information category. It also filters inconsistent information and then serves interesting and important information by using preference for information organization. Therefore, the respondents can dynamically receive information based on their current situation and

preference. This can be compared to reading a newspaper, where the information in OSNs is generally diverse, i.e. news, events, entertainment, etc. The information served by Method 4 will bring the readers up to date without the need for newspapers. Moreover, it helps the respondents to save time in finding interesting information. The benefits of Method 4 are suitable for the lifestyle of Thai respondents, as they usually try to adapt to various situations [109]. Hence, the IFM should allow them to control information on their SNP, independently. The analysis indicated that career and gender impact slightly on the overall results.

Table 5.12: Performance evaluation when considering gender, age, and career for Japanese respondents

Category	Description	Information Overload Solution				Information Filtering Performance				Dynamical Information Feeding				Consistent Information Serving			
		Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4
Gender	Male	2.75±1.28	3.63±0.74	2.50±0.76	3.00±0.76	2.50±1.41	2.63±1.18	2.88±1.46	3.50±1.07	1.88±1.13	3.00±1.07	3.00±1.07	3.63±0.92	3.00±1.41	3.38±1.19	3.25±1.16	3.50±1.2
	Female	2.00±1.10	2.83±0.98	2.33±1.03	2.17±0.98	2.50±1.05	3.00±0.63	3.17±0.75	3.33±1.21	1.83±0.41	3.83±0.41	2.67±1.21	3.17±0.98	2.83±0.75	4.00±0.00	3.67±0.52	3.67±0.82
Age	21-25 years old	2.43±1.51	3.14±1.07	2.43±0.98	2.29±0.76	2.29±0.95	2.57±0.98	3.00±1.29	3.14±1.46	1.86±0.69	3.14±1.07	3.29±0.76	3.71±0.95	2.71±1.11	3.57±0.53	3.86±0.38	4.00±1.00
	26-30 years old	2.33±1.15	3.33±0.58	3.00±0.00	2.67±1.53	3.00±2.00	3.33±0.58	3.33±0.58	3.67±0.58	2.33±1.53	4.00±0.00	1.67±0.58	3.00±1.00	3.00±1.00	4.00±0.00	2.67±0.58	3.00±0.00
	Over 30 years old	2.50±1.00	3.50±1.00	2.00±0.82	3.25±0.50	2.50±1.29	2.75±1.26	2.75±1.50	3.75±0.50	1.5±0.58	3.25±0.96	3.00±1.41	3.25±0.96	3.25±1.5	3.50±1.73	3.25±1.5	3.25±1.26
Career	Student	2.50±1.41	3.13±0.99	2.50±0.93	2.25±1.04	2.50±1.41	2.75±1.04	3.25±1.16	3.50±1.07	2.00±1.07	3.25±1.04	2.88±1.13	3.75±0.89	2.88±1.13	3.63±0.51	3.50±0.76	3.75±1.04
	Engineer	2.33±1.15	3.67±1.15	2.33±0.58	3.33±0.58	2.00±1.00	2.33±1.15	2.00±1.00	3.67±0.58	1.33±0.58	3.00±1.00	2.00±1.00	2.67±0.58	2.33±1.53	3.00±1.73	2.67±1.53	3.33±1.53
	IT specialist	2.33±1.15	3.33±0.58	2.33±1.15	3.00±0.00	3.00±1.00	3.33±0.58	3.33±1.15	3.00±1.73	2.00±0.00	4.00±0.00	3.67±0.58	3.33±1.15	3.67±0.58	4.33±0.58	4.00±0.00	3.33±0.58

Table 5.13: Performance evaluation when considering gender, age, and career for Thai respondents

Category	Description	Information Overload Solution				Information Filtering Performance				Dynamical Information Feeding				Consistent Information Serving			
		Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4	Method1	Method2	Method3	Method4
Gender	Male	2.25±0.71	2.50±0.53	3.13±0.83	3.625±0.92	2.36±0.92	2.36±0.92	2.25±0.71	4.00±0.93	2.36±0.92	2.63±1.06	2.50±0.93	3.25±1.49	2.50±1.07	3.25±0.89	3.375±0.74	4.50±0.53
	Female	2.54±1.33	2.69±1.25	3.23±1.01	3.46±1.05	2.69±1.25	2.85±1.21	3.62±1.19	4.08±1.03	2.85±1.47	3.15±1.14	3.54±0.97	3.69±0.75	3.00±1.47	3.15±1.07	4.46±0.52	4.08±1.04
Age	21-25 years old	2.75±1.26	2.50±0.58	3.25±0.96	3.75±0.96	2.00±1.41	2.00±1.41	2.50±1.00	4.50±0.57	2.25±1.50	2.75±1.5	3.00±1.41	4.00±0.82	2.75±1.70	2.75±1.70	4.00±0.82	4.25±0.50
	26-30 years old	2.45±1.29	2.73±1.27	3.55±0.93	3.45±1.13	2.82±1.25	3.00±0.87	3.55±1.36	4.18±0.87	2.91±1.51	3.09±1.22	3.18±1.17	4.00±0.63	3.09±1.51	3.36±0.80	4.36±0.67	4.45±0.52
	Over 30 years old	2.17±0.75	2.50±0.83	2.50±0.55	3.50±0.84	2.50±0.55	2.50±0.55	2.67±0.82	3.50±1.22	2.50±0.55	2.83±0.75	3.17±0.75	2.33±1.03	2.33±0.52	3.17±0.75	3.50±0.84	3.83±1.47
Career	Student	2.10±1.1	2.50±0.85	2.90±0.74	3.70±0.82	2.10±0.74	2.00±0.47	2.60±0.84	4.10±0.88	2.30±1.06	2.50±0.97	3.10±1.10	3.40±1.17	2.40±0.97	2.70±0.82	4.00±0.94	3.90±1.10
	Engineer	2.83±1.47	2.33±1.03	3.50±1.05	2.83±1.17	2.50±1.38	3.00±1.41	3.50±1.76	3.67±1.36	2.50±1.64	3.00±1.26	2.67±1.03	3.67±1.03	2.83±1.72	3.16±0.75	4.16±0.75	4.67±0.52
	IT specialist	2.60±0.55	3.20±1.30	3.40±1.14	4.00±0.71	3.60±0.89	3.60±0.89	3.60±0.89	4.40±0.55	3.60±0.89	3.80±0.84	3.80±0.84	3.60±1.14	3.60±1.34	4.20±0.84	4±0.71	4.40±0.55

5.6 Discussion

From behavior observation, when they were asked to participate in the experiment, almost all Japanese respondents queried: "When is the deadline?" and answered "はい". This shows the importance of time and style of answer. Some were anxious as they were not sure whether or not they could finish the experiment in time. This behavior is explained in Hofstede's uncertainty avoidance (UAI) dimension [41], referring to the degree of tolerance level when one faces ambiguous situations. Japanese achieve a high UAI score, which indicates that they worry about unexpected circumstances. Hence, they learn to prepare themselves for any eventuality. Although they are worried, they say "はい". In Japanese culture, negative words are considered impolite, so it is necessary to use cues to infer what they really mean. Around 80% of Thai respondents stated: "When I have time, I will do it". Most Thai people favor flexibility. Thai culture is based on the current time or present time. They always make adjustments to suit a person or situation. Time or deadlines in Thailand are changeable [41]. For example, in business, parties may not be able to adhere to an exact deadline in negotiations [122] since Thai business people prefer long-term business arrangements to obtain long-term benefit.

From text analysis, using emoticons and length of text were found to be quite different between Japanese and Thais. This indicates that Japanese and Thai readers will consume different characteristics of text. When using emoticons, most Japanese respondents are likely to use character-based and graphic emoticons in order to express current moods or intentions, such as happiness (^ v ^), greeting ((^_^)/), sleepiness ((。−ϖ−)zzz), and so on. Using emoticons as cues help the readers to interpret the emotional intentions of the creator [123], in particular, where the creator from a high context culture, uses implicit or ambiguous statements in text. The character-based emoticon is more commonly used than the graphic one, which corresponds to the research of Kavanagh et al [124]. This emoticon from text analysis is mixed within the sentences, especially by placing at the end of a sentence. Meanwhile, emoticons are not often used within sentences in Thailand. For the length of text, Japanese creators are more likely to write long sentences to describe their updated status or experiences. On the other hand, Thai creators tend to compose short sentences. This is consistent with the reading behavior in Thai culture, where Thais

prefer to read short sentences. Therefore, the length of text is influenced by culture. The Japanese posting style might not be satisfactory for Thai readers.

There are many advantages to the proposed IFM as follows:

1. The proposed IFM uses the data analysis from Chapter 4, which is not only good for the reader but also for OSN marketing purposes. An advertisement could be shown on the SNP at the best time for promoting the product without leading to the annoyance of its reader.

2. The reader's profile, e.g. current situation, age, and career is used as the input for information filtering, and hence the result will rely on the personality of the reader. Using the reader's profile is an easy way to filter uninteresting information since it does not use other readers' ratings to find similarities between readers and then provide suggestions.

3. The information filtering compares three classification algorithms: DT, KNN, and NB. These results are more reliable in filtering uninteresting information. In the classification, there are many algorithms that are powerful and less time consuming. Nonetheless, if the training set is not suitable for that algorithm, the results are not reliable. This research uses the NB algorithm, which is effective and less complex for filtering uninteresting information. No cold start problem occurs because the content-based approach is applied. Thus, new information, based on the training set, can provide the results quickly.

4. Although this research compares the cultural differences between Japanese and Thais, the analysis results are more general. The proposed IFM can be applied to other cultures or societies with similar characteristics to Thais, such as lifestyle, business negotiation, or working practices. People in Southeast Asia share cultural traits, social freedom, and climate. For example, Thailand and Vietnam are similar in business negotiation. The proposed IFM is also suitable for readers with various daily activity periods, such as engineers or business people. Meanwhile, the analysis results obtained from Japanese respondents can be employed in China, Korea, or other countries, where the cultures are similar to Japanese.

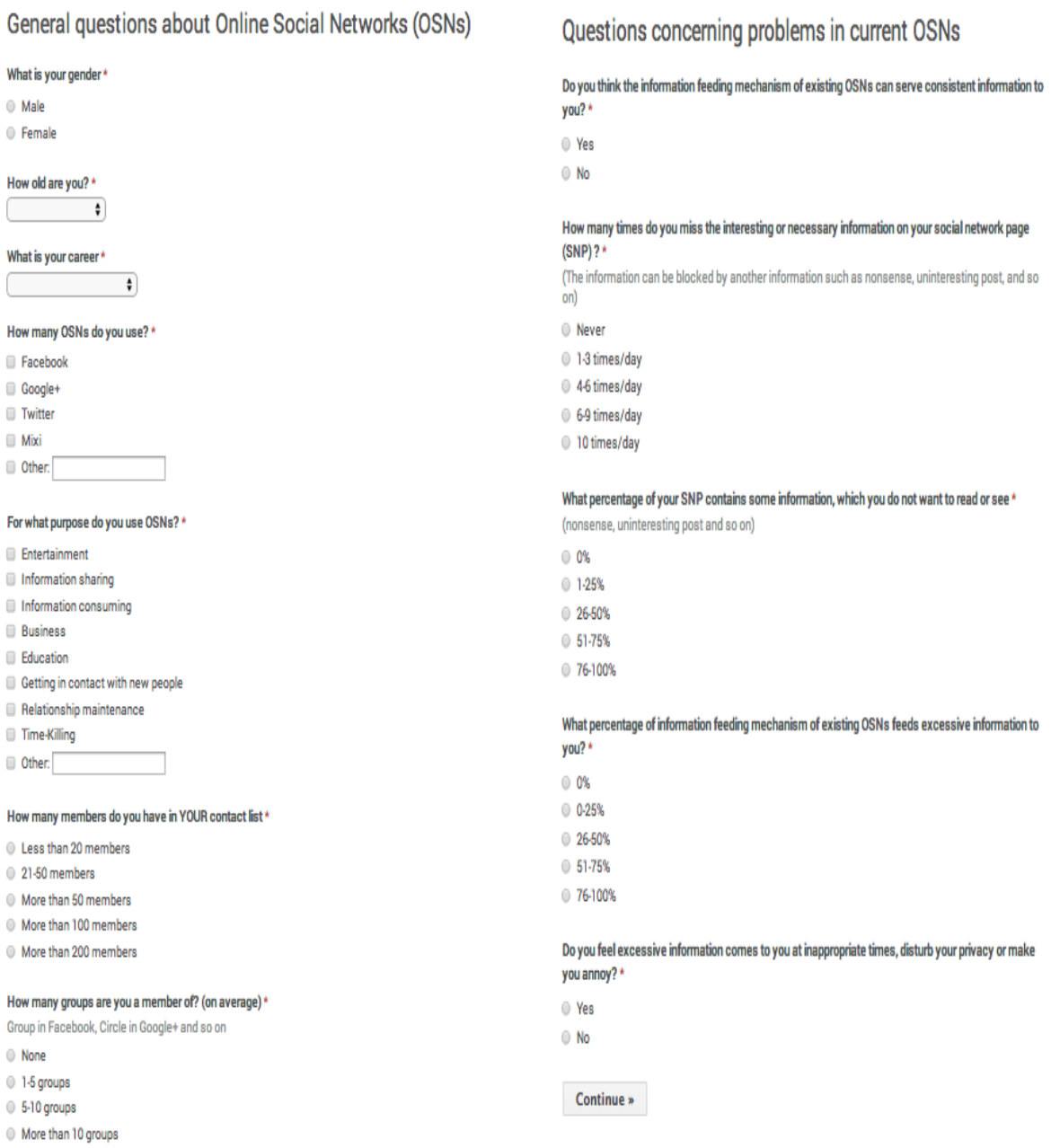
Nonetheless, the proposed IFM has some disadvantages as described below:

1. The proposed IFM might not be inconvenient for the readers because it requires personal information, such as the reader's current situation, age, preference, and so on. However, this personal information will enable the reader to receive information consistent with their requirements.

2. The information category used to filter uninteresting information in the proposed IFM requires high classification accuracy. If the information category is incorrect, this might lead to false predictions in the proposed IFM. For instance, if the actual information category is a personal story, but the text classification predicts it as an advertisement, the proposed IFM misunderstands and filters the personal story. Therefore, this interesting information is missed.

5.7 Conclusion

Several algorithms, sets of features and factors are investigated to reduce information overload to support cultural differences in OSNs. The classification accuracy of the proposed IFM is not very high, but in practice it is more reliable. There are four reasons to describe why the proposed IFM is more reliable. Firstly, the classification performance can be more accurate by improving text classification and using interesting levels of information (reduction of false positive and false negative), as described in Chapter 8. Secondly, the set of features and factors relates to the cultural differences in OSNs. Thirdly, the NB algorithm is less complex, fast, and provides highly scalable model building. Finally, the NB algorithm can be applied to a very large data set in OSNs using the principle of parallelism [118]. The proposed IFM controls the most suitable information based on the reader's current situation and nationality. The reader can dynamically consume interesting and important information in a short space of time based on their current situation. This information is ordered by the reader's preference. The proposed IFM is good for the reader and marketer in OSNs because the advertisement is shown on the SNP at the best possible time for promoting the product with-out leading to the annoyance of the reader. The proposed IFM can be more generally applied to other cultures and societies, even though this study only relates to Japanese and Thai cultures.



(a) Questionnaire page 1

(b) Questionnaire page 2

Figure 5.16: Pages 1 and 2 of the questionnaire used for evaluating the information filtering mechanism

Question about Information feeding method 1-4

In the experiment, what was the current situation in question 3 did you select? *

- Meeting with co-workers <-> Seminar
- Working involving code programming <-> Doing homework or research
- Working involving getting requirement from customer <-> In classroom
- Working involving product analysis <-> Studing
- Private time with family
- Private time with friends
- Private time with alone
- During travel
- Shopping
- Partying

What percentage of each information feeding mechanism solves the information overload problem? *

*Top stories using post's popularity, number of comments, affinity level between reader and creator. **Timeline using reverse chronological order. ***Current situation using user's current situation and other factors (e.g. information category, age, and career)

	0%	1-25%	26-50%	51-75%	76-100%
*Top stories					
**Timeline					
***Current situation					
Random post					

What percentage of each Information Feeding Method contains information overload? *

	0%	1-25%	26-50%	51-75%	76-100%
Information Feeding Method 1					
Information Feeding Method 2					
Information Feeding Method 3					
Information Feeding Method 4					

Do you think feeding the information based on your current situation is useful? *

- Yes
- No

(a) Questionnaire page 3

What percentage do you think feeding the information based on current situations helps you receive the suitable information? *

- 0%
- 1-25%
- 26-50%
- 51-75%
- 76-100%

What percentage of each Information Feeding Method can filter out uninteresting/ unimportant/ nonsense information in your social network page (SNP)? *

	0%	1-25%	26-50%	51-75%	75-100%
Information Feeding Method 1					
Information Feeding Method 2					
Information Feeding Method 3					
Information Feeding Method 4					

What percentage of each Information Feeding Method can dynamically feed the information according to your current situation? *

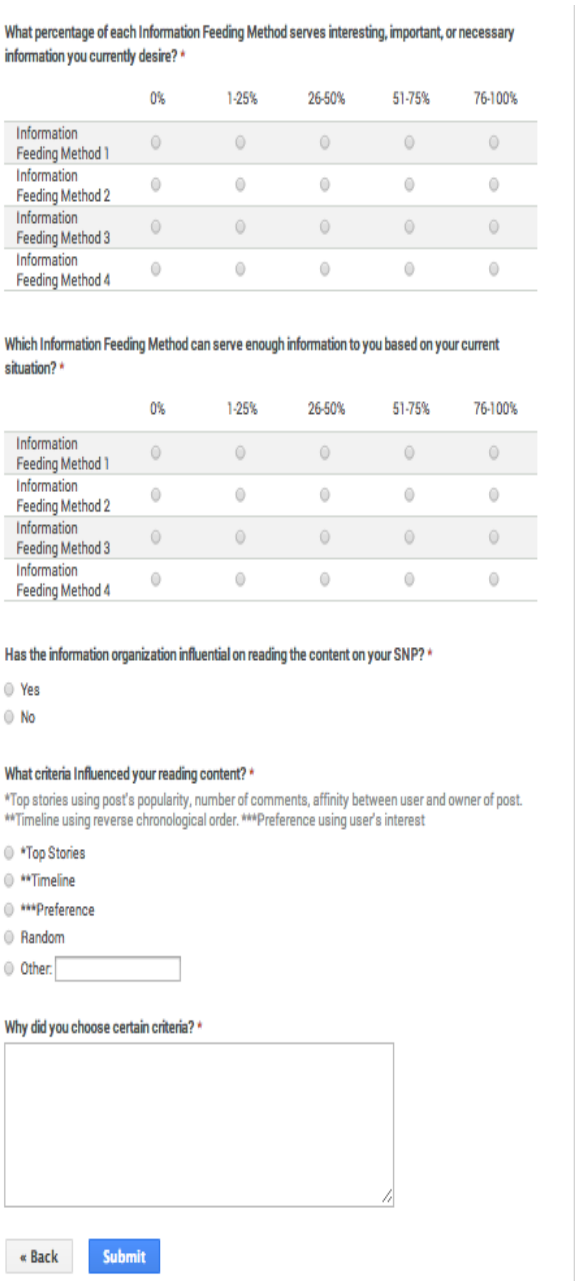
	0%	1-25%	26-50%	51-75%	76-100%
Information Feeding Method 1					
Information Feeding Method 2					
Information Feeding Method 3					
Information Feeding Method 4					

What percentage of each Information Feeding Method serves consistent content to you at current desire? *

	0%	1-25%	26-50%	51-75%	76-100%
Information Feeding Method 1					
Information Feeding Method 2					
Information Feeding Method 3					
Information Feeding Method 4					

(b) Questionnaire page 4

Figure 5.17: Pages 3 and 4 of the questionnaire used for evaluating the information filtering mechanism



(a) Questionnaire page 5

Figure 5.18: Page 5 of the questionnaire used to for evaluating the information filtering mechanism

Chapter 6

Collective privacy protection for information sharing

6.1 Introduction

Generally, the creators in Online Social Networks (OSNs) have the freedom to generate and publish any information to others anywhere at any time via his/her Social Network Page (SNP) as well the SNP of friends. Moreover, he/she can tag and mention other users with such information. However, these actions might cause the creator and multiple users, associated with the information, a loss of privacy due to lack of adequate privacy protection.

6.1.1 User and information definitions

In this research, a user of OSNs can be a reader and creator as shown in Figure 6.1. The reader represents the user, who consumes the information via the SNP. On the other hand, the creator is a general term used for the user, who creates and posts private and collaborative information on OSNs. The private information belongs to only this particular creator, whereas collaborative information, i.e. text, photo, etc., is associated with multiple creators. Owner and co-owner are subsets of the creator. The owner can create and post collaborative information on OSNs. The co-owners can participate or can be referred to by the owner, such as with tagging, mentioning, and sharing collaborative in-

formation. The scenarios below have been obtained from Hu et al. [13] and indicate how the owner and co-owners can be defined. Nonetheless, the user definition in this research is different from Hu and team's research.

A photo is taken by Alice, Bob, and Carol and is collaborative information. If Alice posted this photo via her SNP and tagged Bob and Carol into it, Alice is the owner, while Bob and Carol are the co-owners. In another instance, Alice wrote a note in which Carol is mentioned as *@Carol*, and posted it on Bob's SNP. *"@name"* refers to mention. Alice is the owner because the note was created by her. Bob is also defined as a co-owner because the note is posted on his SNP. Carol is given the position as a co-owner since she was mentioned. Moreover, OSNs allow users to share information, which does not belong to them. For example, Alice saw and shared the interesting photo appearing on Bob's SNP. Bob is known as the creator of this photo.

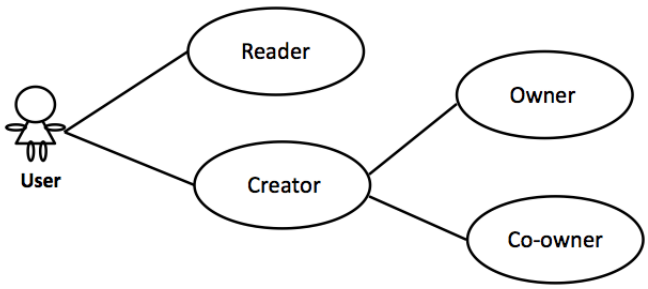


Figure 6.1: Definition of a user in OSNs

Besides private and collaborative information, this research defines additional types of information, which are general information and sensitive information. General information refers to content not detailed or specific to a person. Sensitive information means content that can be used to identify an individual or group of people. It has a quick and delicate appreciation of others' feelings. Five examples of sensitive information in this research are as follows:

1. Personal information (i.e. location, e-mail, address, phone number, etc.).

2. Confidential business information (i.e. trade secrets, sales and marketing plans, new products, etc.).

3. Freedom of expression (i.e. expressing of political opinions, religious beliefs, royal institutions, etc.).

4. Improper morality (i.e. drinking alcohol, infidelity, lying, etc.).

5. Embarrassing behavior (i.e. nose picking, kissing in public, teasing, etc.).

6.1.2 Problem definition

When the owner posts collaborative information (e.g. text, photo, video, link) into OSNs, this information is frequently viewed as belonging to that particular owner. Only the owner can manage this information, while the co-owners do not have permission to control it and might not realize that their information is being managed by others [23][125]. From the above scenarios, information is easily leaked by tagging, mentioning and sharing to unwanted target readers (with whom the owner and co-owners are not willing to share) [90]. More details of privacy concerns can be found in Chapter 1. At present, these actions are meaningless for privacy protection because co-owners merely acknowledge that they are being tagged or mentioned through their information, as it has already been shared. However, they do not have permission to control the information before it is spread in OSNs. As a result, the owner and co-owners might lose privacy. In the case of an information leak, solving consequential problems with collaborative information is harder than for private information because it affects many co-owners. Therefore, privacy protection is essential.

In the case of a photo, many creators upload it on OSNs because it can say more than words, and indicates atmosphere, emotions, etc. Moreover, a photo can be easily recognized by many readers. Recently, Facebook mentioned uploading 2.5 billion photos each week [21]. When a photo is taken by Alice (the owner in this research), Alice has rights to control copying, adaptation, publication, and so on. However, Bob and Carol (the co-owners in this research) appear in the photo, and therefore they have rights to portrait as illustrated in Figure 6.2. There is no restriction when taking a photo, but the owner has to ask permission from the co-owners when he/she wants to post this photo in the public domain. This can avoid violation of the co-owner's privacy.

Figure 6.2: Copyright and portrait rights in the photo

Moreover, posting collaborative information might lead to a crime problem because as well as text, photos, videos, and links, OSNs allow the owner to use the "Check-in" feature. This feature can reveal the actual location where certain activities are being performed or carried out by the owner and co-owners. In this sense, a criminal can take advantage of such information. The criminal takes some time to find a victim's location, which is available on the SNP. However, if the criminal already knew the location due to having a close relationship with the victim, posting collaborative information makes the criminal aware of the victim's current activities.

For instance, Noonee Cool (account name) and her family take a nice trip to Japan. She updates her status and tags her sister as shown in Figure 6.3. This makes the criminal realize that Noonee Cool's family is not at home. If the criminal has a close relationship with some of people in the photo, it is possible that the criminal knows their home location and may steal valuables. Therefore, using OSNs without taking care with privacy could become a big problem.

To address this problem, this research proposes Collective Privacy Protection (CPP) to balance the need for privacy protection with information sharing. It enables the owner to create the privacy policy and co-owners to make a decision by voting on it. This can identify and reduce privacy conflicts because at least one co-owner intends it to be kept private. In other words, privacy conflict arises out of different privacy concerns over

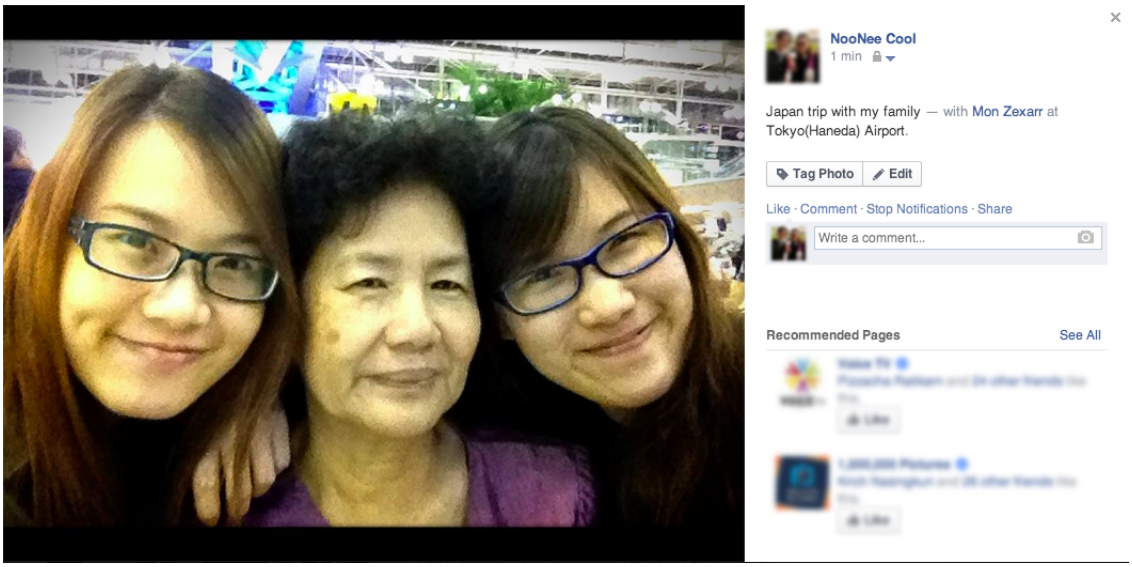


Figure 6.3: The crime problem in OSNs

collaborative information by the owner and co-owners. This research protects privacy covering the owner and co-owners. The collaborative information will not leak to unwanted target readers by implementation of a maximum boundary. Furthermore, when co- owners want to share information, he/she also can portray him/herself as the owner, and create the privacy policy. This is because the owner and co-owners have rights in the collaborative information. It is impossible to make a single privacy policy suitable for everyone because each owner might have different privacy preferences.

6.2 Architecture and methodology

In order to protect the privacy of the owner and co-owners, this architecture as depicted in Figure 6.4 and prevents information from leaking to unwanted target readers. This is because when the owner wants to post collaborative information on OSNs, the co-owners will be notified and vote on the privacy policy created by the owner. This architecture also reduces problems with robbery, kidnapping, and so on. The proposed CPP comprises five main components: social graph, privacy policy, co-owner invitation, majority vote, and conflict identification and solution. More details of each component are described in the next Section.

The work-flow of the proposed CPP begins when the owner creates the privacy policy. Then, the co-owner is detected in order to send an invitation that he/she owns a part of the information, and it is being posted to other readers in OSNs. When co-owners receive the invitation and read the privacy policy, he/she can vote on it. The co-owner can indicate one of three statuses: acceptance, rejection, and no response. Nevertheless, when the time limit ends, the co-owners with no response may move the status to rejection because privacy is considered as a high priority. Next, the proposed CPP finds privacy conflict among the owner and co-owners and provides a solution for each conflict. A list of target readers who can see this information, based on the privacy policy is suggested to the owner to re-check before posting it onto OSNs.

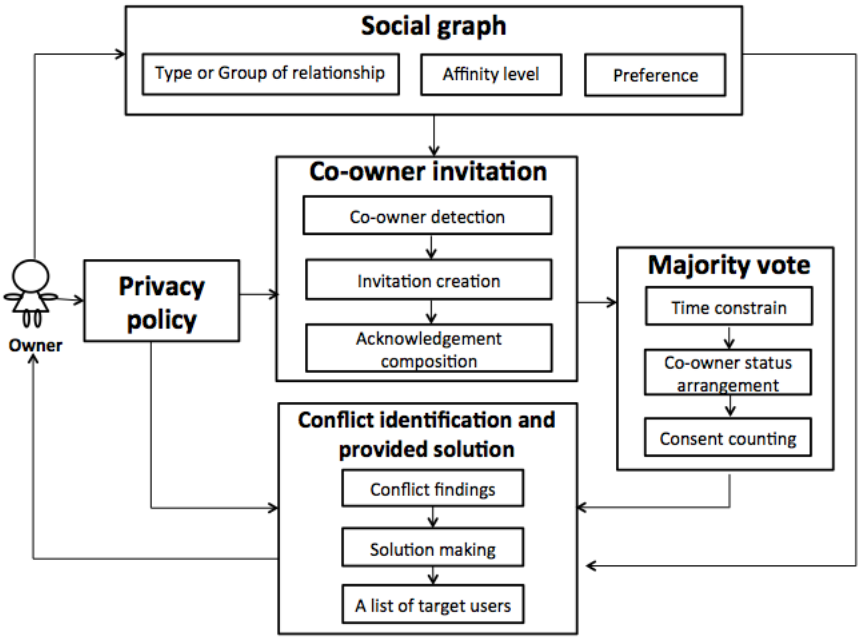


Figure 6.4: The proposed CPP for the owner and co-owners in information sharing

6.3 The proposed collective privacy protection

The proposed CPP comprises five main components: social graph, privacy policy, co-owner invitation, majority vote, and conflict identification and solution. The photo in this research focuses on one in which the co-owner's face can be clearly seen, and where there are few co-owners with portrait rights. The photo taken of an event in a public

place contains an enormous amount of people attending the event and indicates certain facts and atmosphere at that time. It is possible that a person's face might be so small so as not to be recognized, or it might be shown sideways. Therefore, people in this kind of photo are not considered to have portrait rights.

1. **Social graph**

   This study aims to create a graph that represents the social relationship among the users in OSNs, as demonstrated in Figure 6.5. A node refers to a user, such as a Facebook user, Googel+ user, and so on. An edge presents the relationship between two nodes. The label between nodes indicates the type or group of relationship and its affinity level. In this research, preference of the user is also added to the social graph. This graph supports the notion concerning the importance of relationship quality [126] because relationships between users influence privacy decisions.

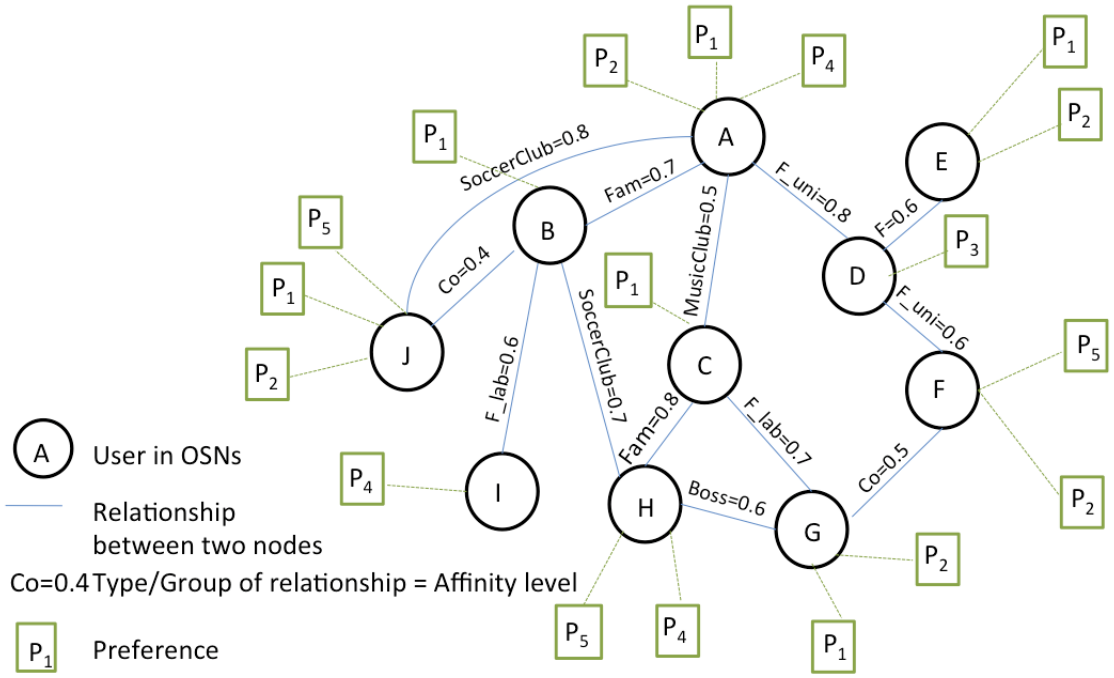


Figure 6.5: A simple social graph in OSNs

2. **Privacy policy**

   The privacy policy is designed to limit the number of the readers, who can see the collaborative information. The idea of the privacy policy is for the owner to

try and match this information to the target readers, who might be interested in it. Nonetheless, when co-owners want to share collaborative information, they can change their position to that of the owner and can then create the privacy policy. Using the privacy policy allows the owner to know who can see the information. At the same time, it helps to protect privacy because collaborative information cannot leak to unwanted target readers. Creating the privacy policy is flexible depending on the purpose for sharing information (both private and collaborative information). The owner constructs the privacy policy by using four useful factors obtained from a social graph: type or group of relationship, distance of information distribution, affinity between the owner and target readers, and preference of target readers. The privacy policy helps to alleviate information overload by reducing the amount of information. The owner can control the distance of the information spread in OSNs.

- **Type or group of relationship** (T/G_Rela)

  In OSNs, the users, both creators and readers, have the ability to create a group for different purposes, such as Group in Facebook, List in Twitter, and Circle in Google+. It is a fact that members in a contact list cannot have the same role in both the world of OSNs and the real-world. Therefore, type or group of relationship is subject to each creator. For example, if the owner does not want other readers, other than a group of friends from university, to see the collaborative information, it can be customized by the owner using type or group of relationship.

  T/G_Rela = {F, Fam, B, Co, T,...} where F=friend, Fam=family, B=Boss, Co=Co-worker, T=Teacher.

- **Affinity level** (AL)

  This factor refers to the familiarity of two users or how often they interact with each other. Generally, the creator is not able to give everyone in the contact list the same level of closeness. For example, if there are 20 members in a friend group from the same university, the creator cannot interact or familiarize themselves with everyone equally.

  AL = {0.1, 0.2, 0.3, ..., 1.0} where 0.1 is very unfriendly, 1.0 is very familiar.

- **Preference** (Pref)

  The user's preference presents their interests, such as music, movies, sport, and so on. This is a good method to use because information will not be posted to those readers who are not interested, so it does not annoy them.

  Pref = {P_1, P_1, P_3, ... , P_k } where P means music, TV, sport, politics.

- **Distance for information distribution** (Dist)

  This indicates how far the information can be spread to other readers. Controlling the distance can limit the number of the readers, who can see the information. Nevertheless, it relies on the purpose of the owner. In other words, if the owner intends to spread information as much as possible without using a privacy setting, it is possible that such information will be consumed by a large number of readers. For example, the default distance for information distribution in Facebook is 1 hop. This means that all members in the contact list can see this information and can share it with their friends. Generally, the information in OSNs can be distributed within 2 to 5 hops.

  Dist = The depth in the social graph from one node to another.

3. **Co-owner invitation**

   This component is proposed to inform co-owners that they are part of the collaborative information. It differs from previous works [85][90] because this component allows co-owners to know that their information is being managed by others. In many cases, co-owners lose privacy caused by the owner sharing collaborative information without permission. For example, three persons (the owner and co-owners) appear in a photo taken during a drinks party. This photo is posted by the owner using a photo file. This photo is then seen by the family of the co-owner, who is obviously concerned about their feelings. Additionally, this component is important since it helps co-owners to realize whether or not posting collaborative information might create problems for them in the future. In this sense, the co-owner can make a decision by voting on the privacy policy created by the owner. This component consists of three sub-components: co-owner detection, invitation creation, and acknowledgement composition.

- **Co-owner detection**

  This means locating the associated co-owners. At present, there are many tools for detecting co-owners. For example, Facebook provides photo face detection.

- **Invitation creation**

  This is used for sending out the privacy policy created by the owner through OSNs or e-mail (if the co-owners do not use OSNs) to the detected co-owners. When co-owners receive the invitation, they can vote or make a decision whether or not to allow the information to be posted to OSNs. Co-owners can consider the privacy policy and estimate the number of readers, who have a relationship with them.

- **Acknowledgment composition**

  After the co-owners vote on the privacy policy, the result is then collected and transferred to the majority vote component. The voting result of each co-owner is considered to be the co-owner's status. The co-owner can choose one of three statuses: acceptance, rejection, and no response. Anacceptance status means the co-owner agrees with the privacy policy. A rejection status indicates the co-owner denies the privacy policy, or he/she needs to keep this information private. A no response status represents that the co-owner does not accept or reject within the allotted time.

4. **Majority vote**

   This component has a duty to seek the consent of all co-owners as much as possible because allowing owners and co-owners to create the privacy policy makes it difficult to meet all desires at one time. It starts with gathering the status of all co-owners from the co-owner invitation component. However, it will take time to collect the voting results from all co-owners, so a specific time needs to be allotted. The co-owner providing a no response status is then moved to a rejection status to protect privacy, when the time limit is reached. After status arrangement, the vote results are then counted and if the number of acceptances represents more than half the results of the vote, this means that the collaborative information can be posted to OSNs. The advantage of the majority vote is that if one co-owner rejects this

privacy policy, he/she is still provided with privacy protection.

5. **Conflict identification and provided solution**

   This component is designed for finding conflicts among those co-owners who accept and reject the privacy policy and creating a solution for those conflicts. Thereafter, a list of target readers is provided to the owner. The owner can verify the information before uploading it. The conflicts are also identified when sharing occurs. In order to find conflicts, the social graph is required because it can indicate how the association or connection of each user. Moreover, it can represent mutual friends. For example, in Figure 6.5, user C is a mutual friend of users A and B, while user H is a mutual friend of users B and C. The mutual friend is necessary to detect conflict between those co-owners who accept and reject the privacy policy. Below is an example scenario showing how conflicts occur, and how they are resolved.

   Alice, Bob and Carol took a photo together at a party using Alice's camera. Alice has the photo file and she wants to post this photo on Facebook. In this scenario, the photo is viewed as collaborative information, which belongs to Alice, Bob, and Carol. By using the proposed CPP, five example cases for conflicts can be detected as shown in Tables 6.1, 6.2 and 6.3. The owner in each case is assumed to accept the privacy policy. Then, if the number of acceptances is more than half, the owner can post the collaborative information.

   Case 1: In Table 6.1, Alice is considered to be the owner. Bob and Carol are defined as the co-owners. If Bob accepts the privacy policy created by Alice while Carol rejects it, this photo can be posted since the positive vote results are more than half (2 acceptances: 1 rejection). From Figure 6.6, users A, B and C represent Alice, Bob, and Carol. A summary of conflicts and a solution are provided as follows:

   - *Conflict*
     - Users H and F have a conflict because user C has rejected the privacy policy.
   - *Result*

– Users H and F cannot see the collaborative information (protecting the privacy of user C)

Table 6.1: Example case 1 when the owner is Alice

Case	**Owner Alice**	**Co-owner Bob**	**Co-owner Carol**
1	Privacy Policy	Acceptance	Rejection

In Table 6.2, when Bob wants to share this collaborative information with OSNs, Bob changes his position from co-owner to owner and then creates the privacy policy. There are two possible vote results. Firstly, in case 2, Alice accepts, but Carol rejects the privacy policy. Secondly, in case 3, Alice rejects, but Carol accepts the privacy policy.

Table 6.2: Example cases 2 and 3 when the owner is Bob

Case	**Co-owner Alice**	**Owner Bob**	**Co-owner Carol**
2	Acceptance	Privacy policy	Rejection
3	Rejection		Acceptance

Case 2: Figure 6.7 illustrates that the user I has a relationship with both users B and C, but user C rejects. Therefore, user I is conflicted and cannot see the collaborative information. In addition, users H and F cannot see the information thanks to user C's rejection of the privacy policy created by user A.

- *Conflicts*
  - User I has a conflict because of user C's rejection of the privacy policy created by user B.
  - Users F and H are conflicted because user C has never rejected the privacy policy created by user A (case 1)
- *Result*

- Users F, H, and I cannot see the collaborative information

Case 3: Alice rejects, but Carol accepts the privacy policy created by Bob. Figure 6.8 shows that six conflicts are found with users F, H, J, K, Y, and Z, caused by user A' s rejection. These conflicts can be separated into two groups and summarized as follows:

- *Conflicts*
  - Users F, J, K, Y, and Z are conflicted because user A rejects the privacy policy created by user B.
  - Users H and F are conflicted because user C rejected the privacy policy created by co-owner A (case 1).
- *Results*
  - Users J, K, and Y can see the collaborative information because these users have never seen the information posted by user A.
  - User Z cannot see the collaborative information because user A rejected the privacy policy created by user B.
  - Users F and H cannot see the collaborative information because user C rejected the privacy policy created by user A in case 1.

In Table 6.3, Carol wants to update what is happening in daily life by using the collaborative information. Carol moves her status from co-owner to the owner and creates the privacy policy. The examples in Table 6.3 continue using case 2. Case 4 presents that Alice accepts, but Bob rejects the privacy policy. On the other hand, case 5 shows that Alice rejects, but Bob accepts the privacy policy. Each case is described below.

Case 4: Alice accepts, while Bob rejects the privacy policy created by Carol. Figure 6.9 shows the privacy conflicts in case 4.

- *Conflicts*
  - Users F and I are conflicted because user B rejects the privacy policy created by user C.

Table 6.3: Example cases 4 and 5 when the owner is Carol

Case	Co-owner Alice	Co-owner Bob	Owner Carol
4	Acceptance	Rejection	Privacy policy
5	Rejection	Acceptance	

  - User I has a conflict because user C has never rejected the privacy policy created by Co-owner B (case 2).
  - Users H and F are conflicted because user C has never rejected the privacy policy created by Co-owner A (case 1).

- *Results*
  - Users F and I cannot see the collaborative information because user B rejects the privacy policy created by user C.
  - Users I, H, and F cannot see the information because user C intends not to reveal collaborative information to users I and H.

Case 5: Alice rejects, whereas Bob accepts the privacy policy created by Carol. The results from case 2 are used to find the privacy conflicts in this case. The social graph in Figure 6.10 indicates the privacy conflicts, providing a list of the users.

- *Conflict*
  - Users H and F since owner C has never rejected the privacy policy created by co-owner A in case 1.
- *Results*
  - User I cannot see the collaborative information although user B accepts the privacy policy created by user C because user I does not pass the privacy policy.
  - Users H and F cannot see the information.

In terms of temporal changes, after posting the collaborative information, if the owner realizes this information causes trouble to him/her, they should have the right to delete

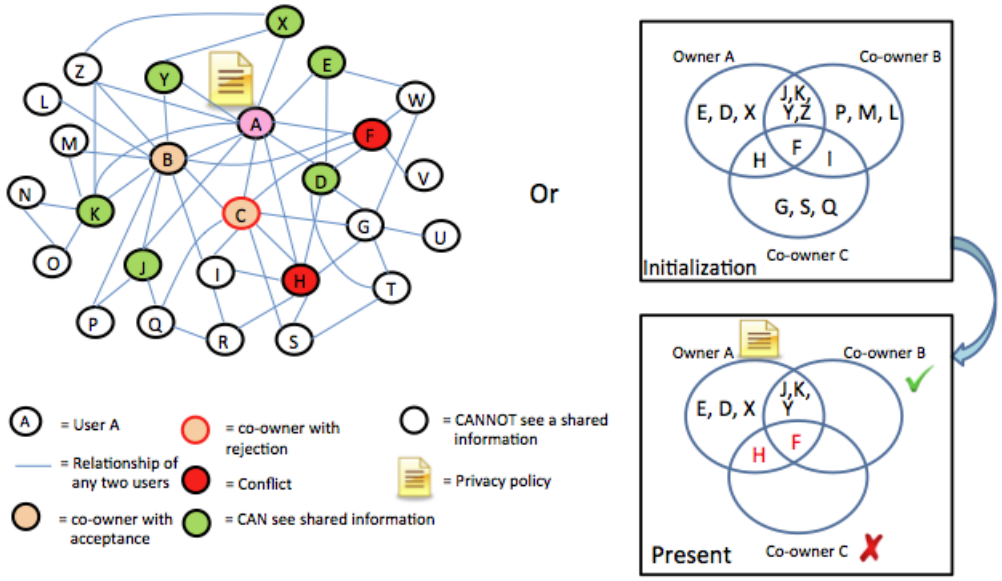


Figure 6.6: The privacy conflict caused by Carol's rejection in case 1

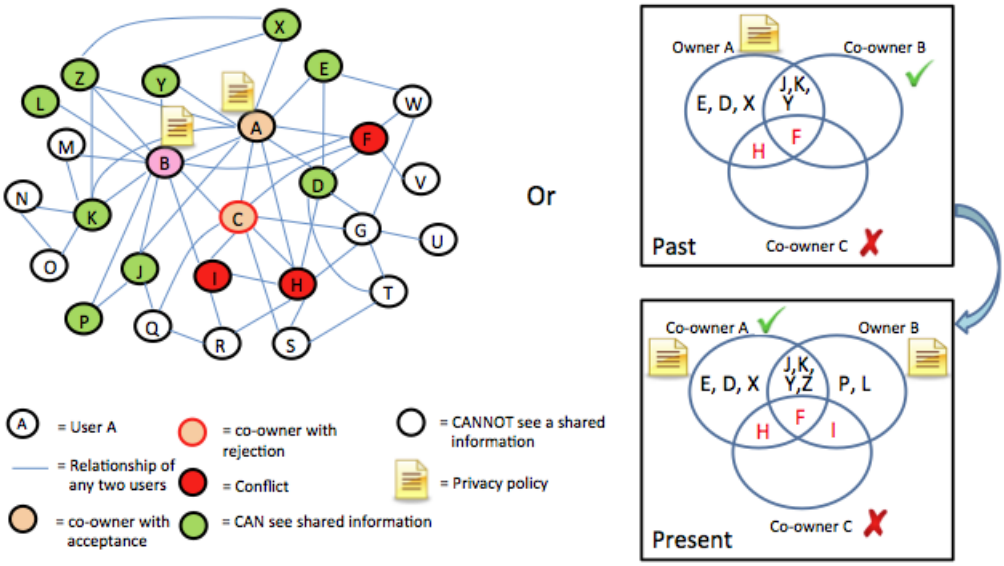


Figure 6.7: The privacy conflict caused by Carol's rejection in case 2

it. Deleting collaborative information can be done without asking permission from co-owners under the policy of the owner, which has never been set. This differs from the process before posting the collaborative information, where the co-owners are asked for permission for privacy protection.

When the social graph changes, the opportunity to see collaborative information is also adjusted, depending on the co-owner's answer at that time. If the co-owner rejects

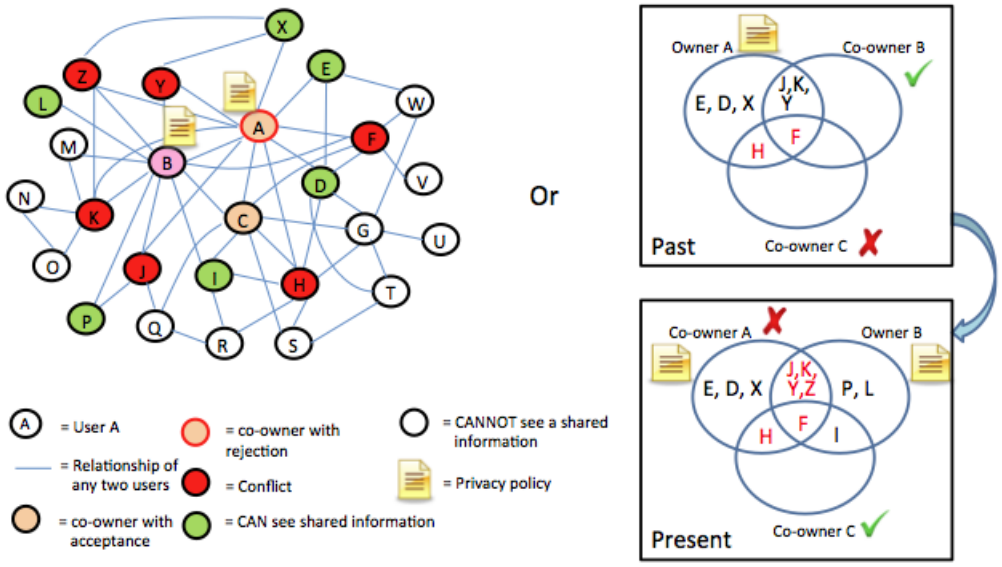


Figure 6.8: The privacy conflict caused by Alice's rejection in case 3

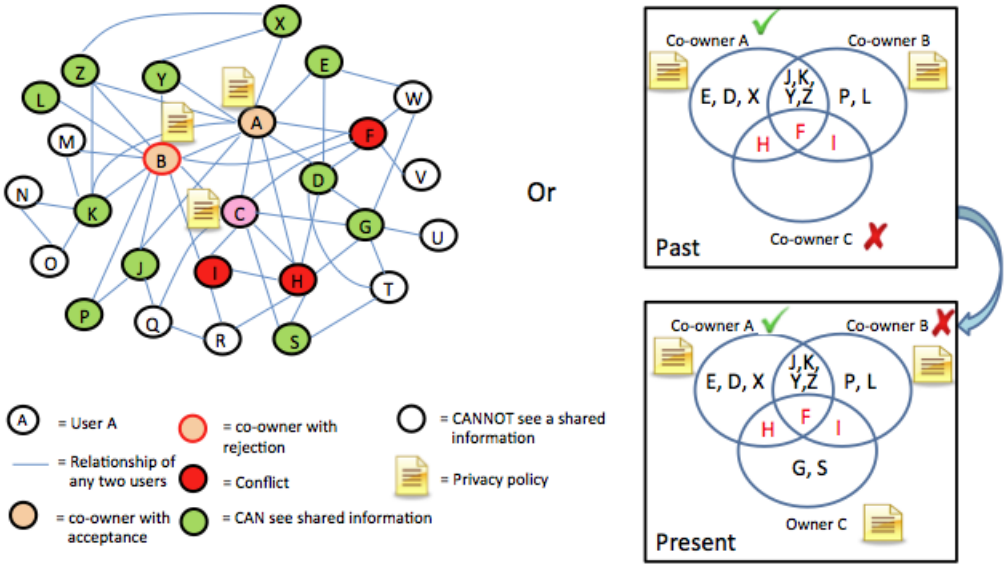


Figure 6.9: The privacy conflict caused by Bob's rejection in case 4

the privacy policy, the reader, who has never seen this information, having just associated with this co-owner, cannot see it. Although this information was seen by many readers, if it disappeared at some point in the future, those readers might not be too surprised. This is because OSNs provide a great deal of information; for example, Facebook users uploaded 2.5 billion photos a week on average in September 2013 [21]. The readers might not realize that some information is missing, and decide to consume new information.

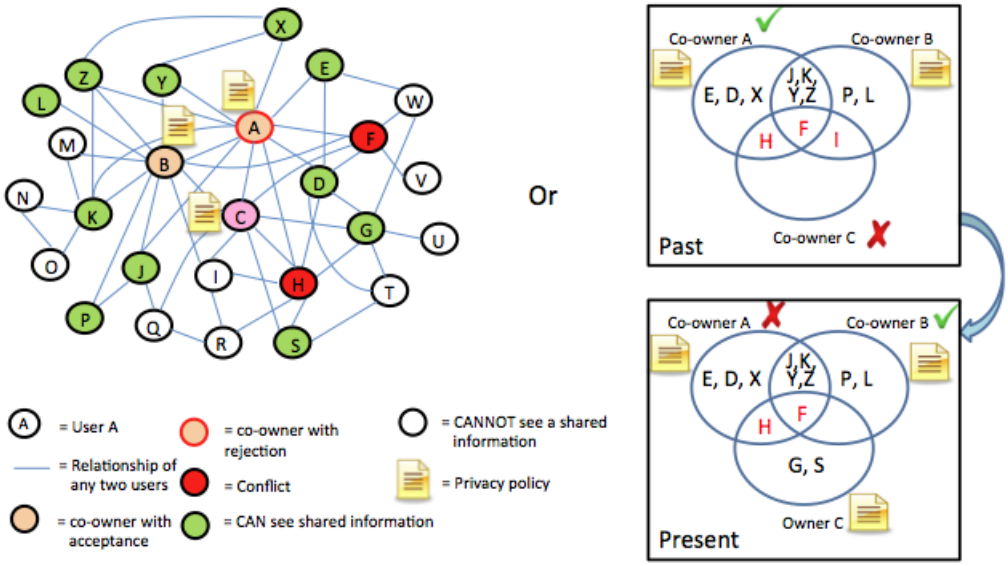


Figure 6.10: The privacy conflict caused by Alice's rejection in case 5

6.4 Experiments and results

This section consists of two experiments, each with different purposes. The first step is to analyze the factor and the combination of factors, which help to prevent information leaking to unwanted target readers. The first experiment also aims to investigate the opinion of co-ownership by using a questionnaire. The analyzed results in this experiment are then used in the privacy policy, which is part of the proposed CPP. The objective of the second experiment is to measure the performance of the proposed CPP for providing a solution when privacy conflicts occur. The proposed CPP is compared with two existing algorithms: naive and Hu solutions [13].

6.4.1 Experimental setup for factor analysis

- **Experimental setup for factor analysis**

  In order to study the factor and combination of factors influential to sharing sensitive information, a virtual social graph was built to help the respondents imagine the flow of information. In this experiment, the virtual social graph was not created by real data, such as a contact list (presenting the type or group of relationship), affinity level and preference in OSNs. This is because permission is required not only from the respondents but also from all members in the respondent's contact

list. Figure 6.11 shows the virtual social graph consisting of 88 nodes and 201 edges. Each node refers to a user of OSNs, and one user had two or three preferences, such as sport, music, games, food, and travel. The edge represents a relationship and 10 affinity levels between two nodes. Each relationship can be one of five different categories: general friend, university friend, family and relatives, co-worker, boss, or teacher. The affinity level denotes the familiarity between two users or how often they interacted with each other. The value of the affinity level ranges from 0.1 (very unfriendly) to 1.0 (very familiar). Meanwhile, one node can consist of more than one edge. In this experiment, the collaborative information was assumed to be associated with one owner and five co-owners. Three co-owners accepted the privacy policy created by the owner, so the vote results showed a majority. There are 15 types for investigation of factor and combinations of factors as follows in Table 6.4.

Table 6.4: 15 types for factor analysis

Number of factor	**Factor and combination**
One factor	1. Type/Group of relationship (T/G_Rela)
	2. Affinity level (AL)
	3. Preference (Pref)
	4. Distance for information distribution (Dist)
Two factors	5. T/G_Rela+AL
	6. T/G_Rela+Pref
	7. T/G_Rela+Dist
	8. AL+Pref
	9. AL+Dist
	10. Pref+Dist
Three factors	11. T/G_Rela+AL+Pref
	12. T/G_Rela+AL+Dist
	13. T/G_Rela+Pref+Dist
	14. AL+Pref+Dist
Four factors	15. T/G_Rela+AL+Pref+Dist

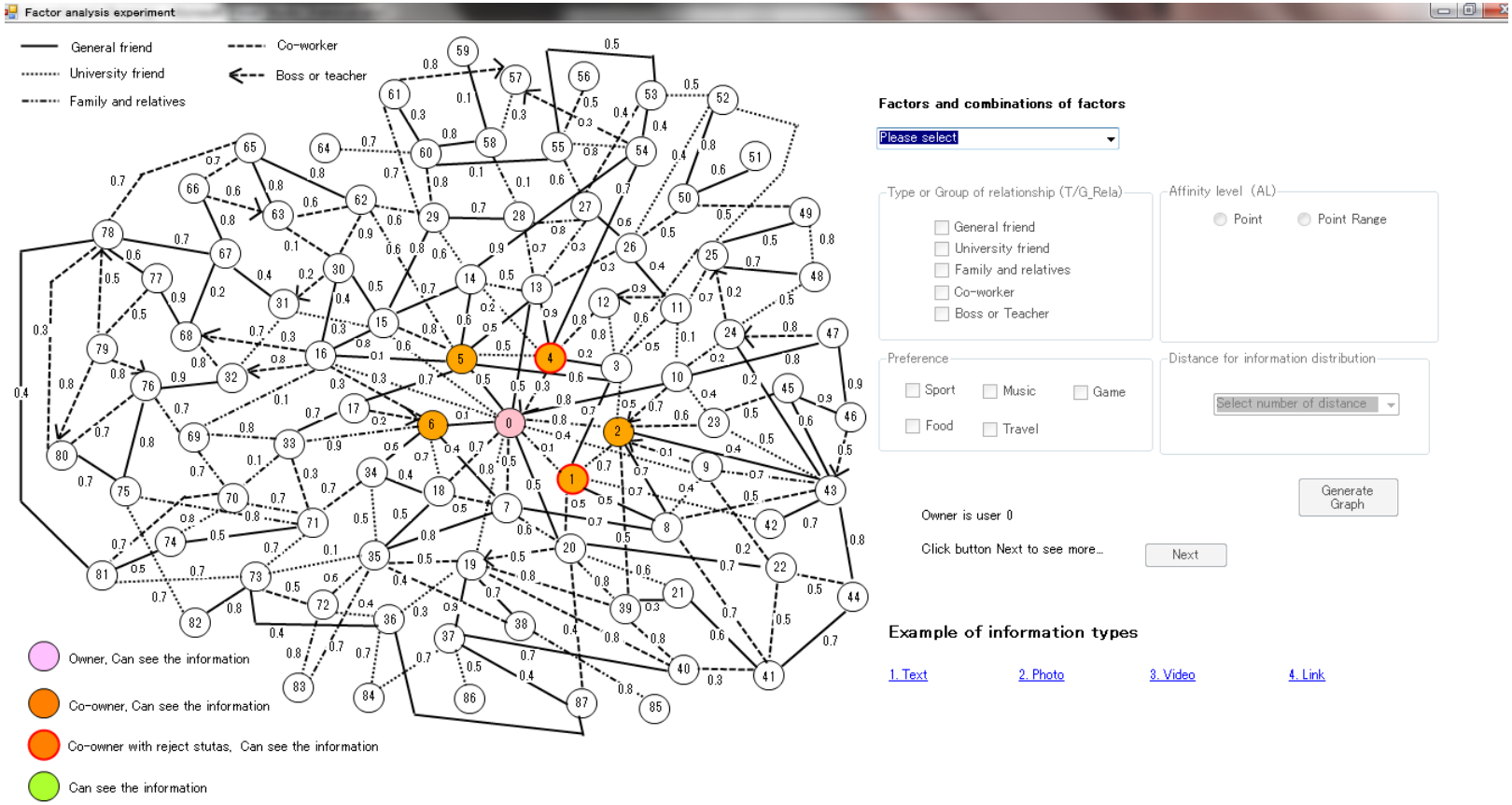


Figure 6.11: An example of the virtual social graph used in the experiment

According to types of information, e.g. text, photo, video, and link. Each respondent is offered many scenarios as denoted in Figure 6.12. The respondent then imagines that they are the owner and create the privacy policy comprising one factor and a combination of factors. They also observe the flow of information in consideration of co-owners, who rejected the privacy policy. In the same privacy policy, the respondent swaps positions with the co-owners and observes the flow of information again.

Figure 6.13 indicates that the owner 0 creates the privacy policy by using the combination of T/G_Rela (setting it as general friend, university friend, and co-worker) and Dist (setting as 1 hop) factors. The co-owners 1 and 4 reject this privacy policy. Therefore, users who have relationships with owner 0 and co-owner 1, and with owner 0 and co-owner 4, cannot see this information since users 20 and 13 are conflicted in this case. Figure 6.14 shows that if co-owner 2 wants to post the collaborative information, their status is changed to owner. Owner 2 also creates the privacy policy by using the same combination of factors but different values (T/G_Rela: co-worker and Dist:2 hops). Nonetheless, co-owners 4 and 6 reject this policy. When considering the policy, users can be divided into two main groups. Firstly, users cannot see the information because of rejection by owner 4 (user 3), which is not consistent with this policy (user 8 and 43). Secondly, users cannot see the information due to acceptance by co-owner 1 (user 20 and 42), which is consistent with this policy (users 9, 10, 11, 22, 23, 24, 39, 45, 47).

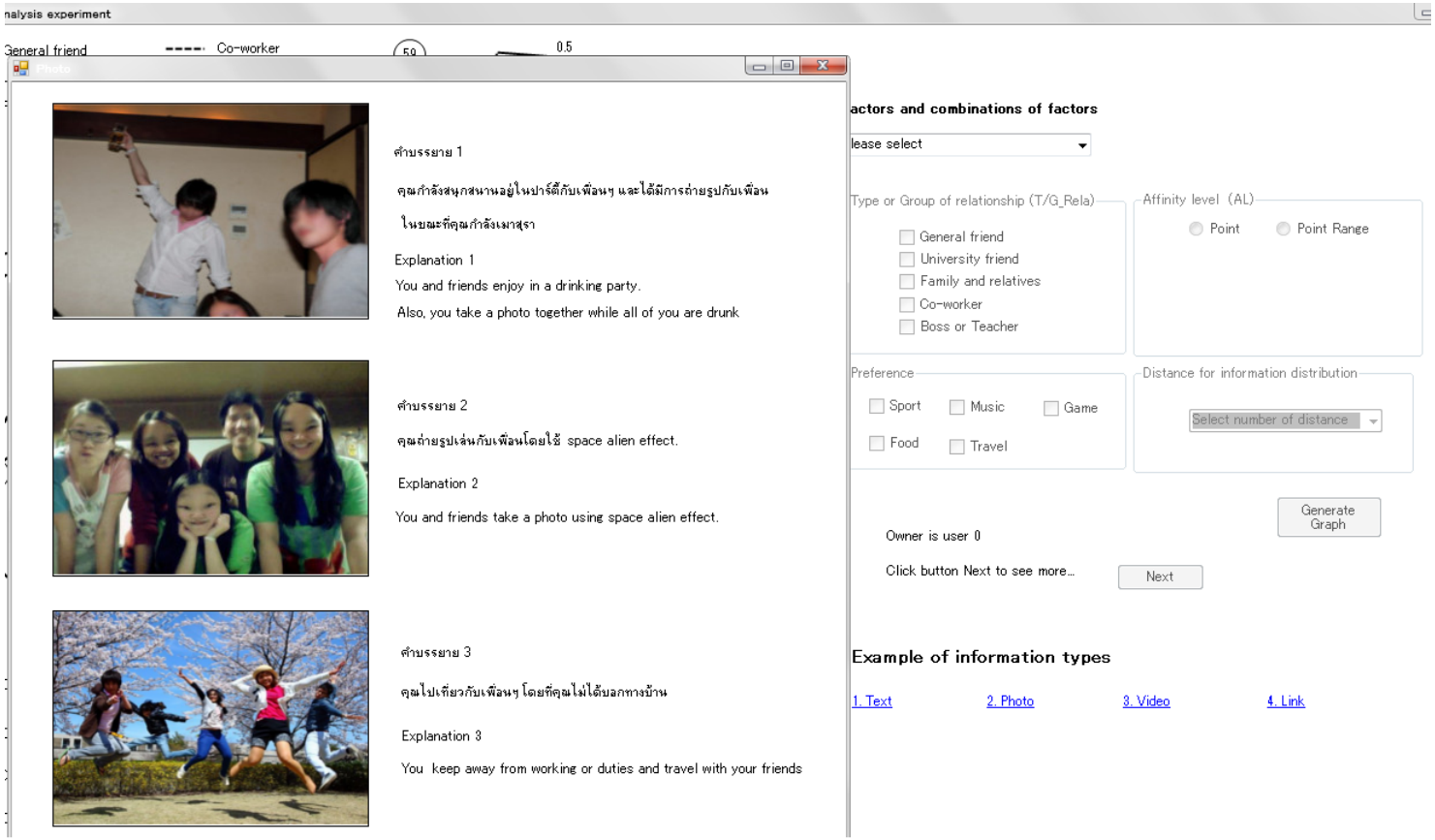


Figure 6.12: An example of scenario

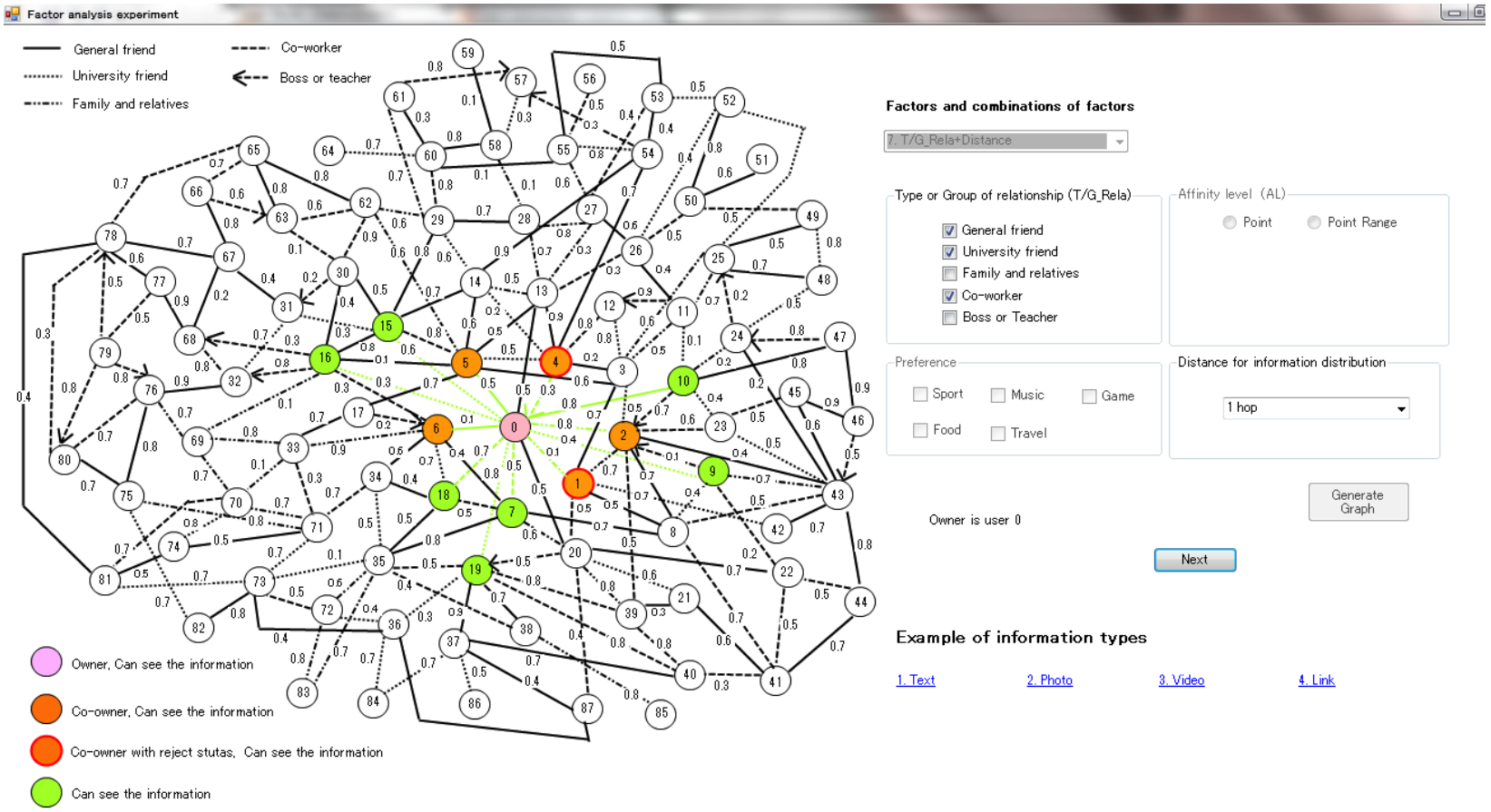


Figure 6.13: Example of changing positions from co-owner to owner and results (before)

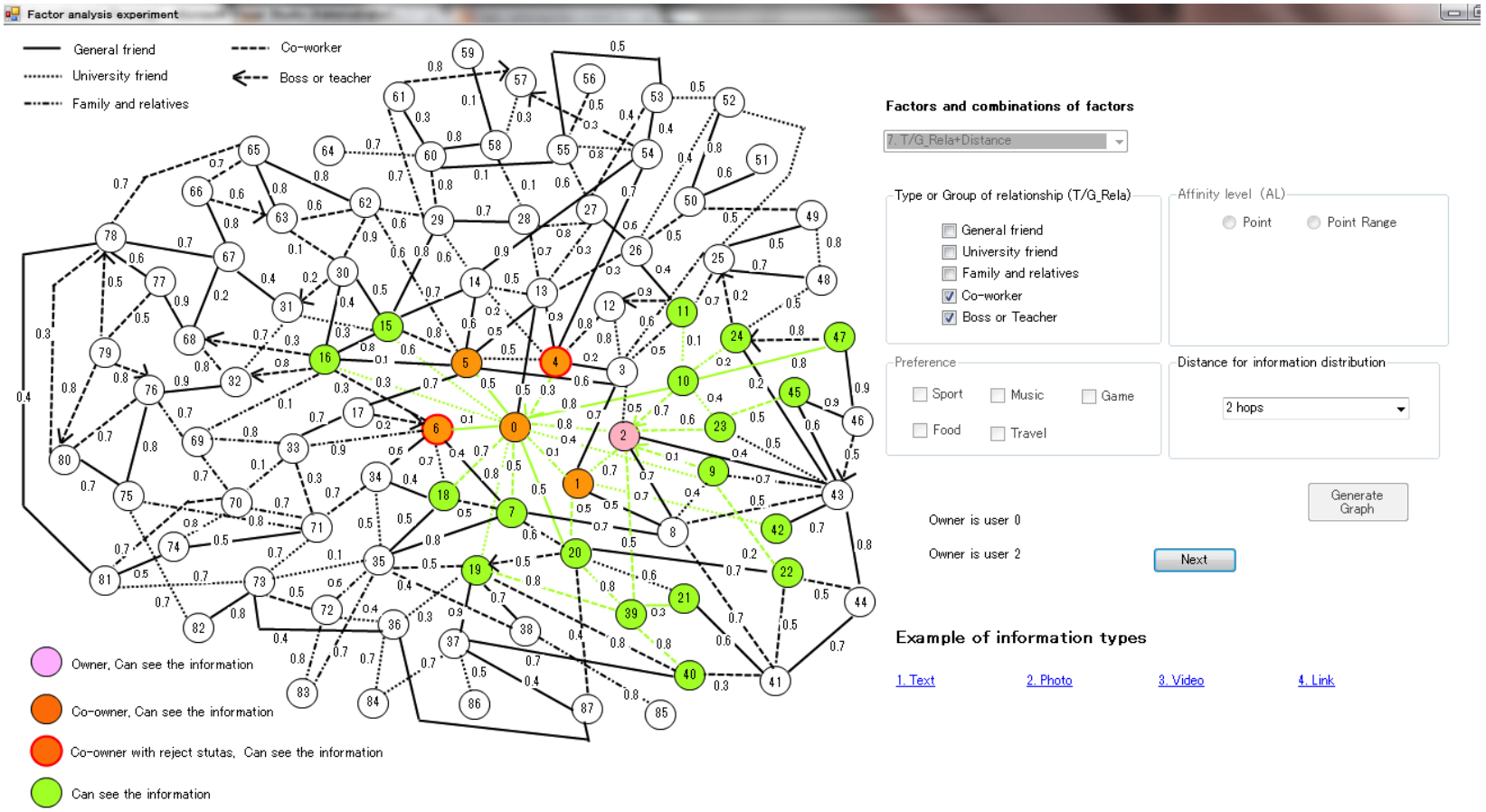


Figure 6.14: Example of changing positions from co-owner to owner and results (after)

After the experiment. The respondents completed the questionnaire to evaluate the performance of each factor and combination of factors. The questions are answered by 24 Thai respondents: 14 male and 10 female. All respondents are asked general questions (i.e. age, gender, use of OSNs, etc.), privacy in OSNs, and opinions concerning co-ownership.

For the general questions, most respondents in the experiment are aged between 21 to 40 years old and have good experience with the use of Facebook, Twitter, Google+, Line (Timeline), and Instagram. They normally have more than one account, i.e. Facebook and Line (Timeline), or Facebook, Twitter, and Instagram etc. 62.5% of respondents spend about 1-4 hours a day on OSNs. The purposes for using OSNs generally relate to entertainment, information sharing, consumerism, business, killing time, and relationship maintenance. Most respondents do not try to make a new friend because their contact list contains more than 200 members, they simply attempt to maintain existing relationships.

- **Results and discussion for factor analysis**

  Each question regarding privacy in OSNs and opinions of co-ownership is answered with Yes/No and a 5-point scale. The performance of each factor and combination of factors in Table 6.4 is investigated using mean and standard deviation.

  1. **Privacy in OSNs**

     - **Privacy setting**

       The result shows that most respondents are concerned with privacy at 3.91 ± 0.99. 62.5% of them make an effort to avoid the leaking of sensitive information by using a custom setting. The remaining respondents use a default setting provided by OSNs. However, they still complain that the privacy setting [91][126][127][128] is difficult to understand as well as guessing the output and do not use the privacy setting. In addition, it is boring because of the many steps required to complete the setting. Some respondents state that they will not post as much information if OSNs provide so little privacy protection. For instance, Line (Timeline) allows

the Line user to select the people who can see the information. They have to perform this step every time before posting.

– **Type of information**

Photo and text are considered to be easily leaked to unwanted target readers. This is because they are easily recognized by others. Furthermore, using tagging and mention on information can be done without any permission. For the types of information such as videos, the level concerning privacy was lower than that of photos and text. Although watching a video helps the respondents to understand the story easily, it can take time to watch an entire video, so they do not often watch them.

– **Concerning group**

The respondents do not want sensitive information to leak to their family and boss. Although the respondents care about them, they still need their privacy and generally perform many roles in society. Thus, they display different behavior when in different social groups. For example, some families do not allow vulgar language to be spoken when in the company of family members and relatives but using such language with friends is acceptable.

Family is the foundation of Thai society, and relationships within the family are closely knit. Thailand has an idiom that "family comes first". Nevertheless, some people are of the opinion that if certain sensitive information is leaked to their family, it can lead to misunderstanding, worry, or disputes.

In circumstances where a boss may have influence over someone's career because certain decisions rely on them, posting information can sometimes affect the image of organization. For example, a teacher's career was ruined because she posted improper activity on MySpace as shown in Figure 6.15.

– **Information deletion and user's behavior**

If information belonging only to the respondent or that which also includes

Teachers' Virtual Lives Conflict With Classroom

May 6, 2008

By SCOTT MICHELS

Like 61 share 32 Tweet 3 +1 0

Stacy Snyder was weeks away from getting her teaching degree when she said her career was derailed by an activity common among many young teachers: posting personal photos on a MySpace page.

Snyder, then 27, claimed in a federal lawsuit scheduled to go to trial Tuesday that Millersville University refused to give her a teaching credential after school administrators learned of a photo on her MySpace page labeled "drunken pirate." She said school officials accused her of promoting underage drinking after seeing the photo, which showed Snyder wearing a pirate hat and drinking out of a yellow cup.

Figure 6.15: Teachers' Virtual Lives Conflict With Classroom [8]

co-owners' leaks to unwanted target readers, most respondents say they would delete it, apologize, and try to offer an explanation to the injured party, especially if it is someone they care about. However, when certain information disappears or is deleted, the respondent in the reader role is not really surprised. He/she understands that the deleted information might lead to a privacy problem for the owner of the information and takes the view that this often occurs in OSNs.

The respondent's behavior in sharing information might influence its leakage. The results report that if the respondent wants to up-date their status in daily life, he/she will post immediately and sometimes other users are tagged or mentioned in that information. The respondents will not be worried if sensitive information leaks to people with whom the respondents do not have a relationship [83][84].

– **Factor analysis**

According to the questionnaire, a number of factors were found to influence privacy protection. In Table 6.5, if the number of factors increases, it helps to prevent sensitive information from leaking to unwanted target readers. Table 6.6 shows the resulting evidence. A combination of T/G_Rela, AL, Pref, and Dist is considered to be most important because it covers the

privacy protection that can filter unwanted target readers. From the overall performance, most respondents do not think that the Pref factor can help protect privacy, thus the performance of the Pref factor or combinations containing the Pref factor are dropped. However, a minority of the respondents reason that combinations with the Pref factor may reduce the scope of certain users, who have similar points of view such as freedom of expression or improper morality.

Taking the respondents' opinions, the T/G_Rela and the AL factors form the preferred basis for privacy setting. Besides these two factors, others may strongly increase privacy protection. This is consistent with the results in Table 6.5. If the combination of factors contains the T/G_Rela and the AL factors, performance will be increased.

Table 6.5: Influence of factors on privacy protection

Factor	Personal information	Confidential business information	Freedom of expression	Improper morality	Embarrassing behavior
1. T/G_Rela	3.54±1.47	3.29±1.60	3.25±1.50	3.75±1.92	3.46±2.00
2. AL	3.42±1.50	2.92±1.53	3.29±1.30	3.46±1.50	3.33±1.34
3. Pref	2.63±1.73	2.67±1.61	3.63±1.38	2.75±1.39	2.54±1.25
4. Dist	3.63±1.12	2.88±1.56	2.92±1.05	3.08±1.21	3.04±1.23
5. T/G_Rela+AL	4.04±1.30	3.58±1.47	3.75±0.98	4.13±1.03	3.71±1.12
6. T/G_Rela+Pref	3.46±1.47	3.33±1.43	3.88+1.08	3.50±1.06	3.17±0.96
7. T/G_Rela+Dist	3.88±1.15	3.38±1.50	3.45±1.30	3.42±1.28	3.29±1.33
8. AL+Pref	3.30±1.31	2.98±1.43	3.83±1.12	3.29±1.23	3.13±1.26
9. AL+Dist	3.46±1.22	3.08±1.50	3.29±1.27	3.38±1.20	3.33±1.17
10. Pref+Dist	3.25±1.22	2.92±1.50	3.25±1.29	2.83±1.17	2.79±1.10
11. T/G_Rela+AL+Pref	3.92±1.32	3.63±1.44	4.08±0.97	4.17±1.00	3.96±1.00
12. T/G_Rela+AL+Dist	4.25±0.85	3.71±1.30	3.79±1.14	4.04±0.95	3.95±1.04
13. T/G_Rela+Pref+Dist	3.83+1.12	3.46±1.38	3.92±1.18	3.92±0.88	3.79±0.78
14. AL+Pref+Dist	3.67±0.96	3.50±1.35	3.96±1.20	3.77±0.95	3.58±1.06
15. T/G_Rela+AL+Pref+Dist	4.50±0.93	3.79±1.44	4.21±1.22	4.38±0.71	4.21±0.72

Table 6.6: Performance of each combination when considering the number of factors

Number of factors	Performance
1	3.17±1.50
2	3.41±1.28
3	3.84±1.11
4	4.22±0.93

2. **Opinion of co-ownership**

   From the questionnaire, it can be noted that most respondents have experienced owners posting collaborative information without their permission. Around 64% of these respondents faced trouble after collaborative information of a sensitive nature was posted to OSNs and worry about information leaks at 3.86 ± 0.77.

   There are two different opinions when the respondents are asked "Should the owner ask the co-owner's permission before posting the collaborative information?". The respondents try to imagine that they are the co-owner. 83.33% of them thought that asking permission was necessary for four reasons:

   - The respondents do not want the information to leak to others with whom they do not want to share.
   - They should have the right to decide whether or not this information can be posted because they cannot know which information will cause them trouble in the future.
   - The sensitivity level of privacy towards each piece of information relies on the individual. In other words, each person has different privacy concerns when seeing the same information. For instance, the owner posts a photo in the OSN. The owner thinks this photo is normal, but some co-owners feel that they have a strange posture. They are embarrassed to share this photo with others. Therefore, asking permission is the proper way.
   - They should know their information is being managed by someone because they are worried who may see it.

In the second group, 16.67% of the respondents state that asking their permission is unnecessary when they are the co-owner. Three reasons are explained below.

- They cannot expect the owner to use the privacy setting, thus the co-owner must themselves be careful with the collaborative information.
- Giving permission every time is a boring task.
- They do not care much about privacy.

In summary, the respondents are worried if collaborative information of a sensitive nature leaks to unwanted target readers. Although they want to post it to OSNs, they need privacy by not revealing some information to others because of negative feedback. The respondents believe that the combination of T/G_Rela, AL, Pref, and Dist factors offers privacy and helps to protect against information leakage. It is expressed that the owner has to take responsibility for asking the co-owners' permission. This is a suitable way forward, although sometimes waiting for permission might mean that the information is not fresh or up to date.

6.4.2 Experiment and results for measuring the performance of the proposed CPP

- **Experimental setup for measuring the performance of the proposed CPP**

The objective of this experiment is to evaluate the performance of the proposed CPP for providing a solution when privacy conflicts occur. Figure 6.16 shows a Venn diagram, where a set refers to all members in a contact list of a user in OSNs, referring to the owner and co-owners. Overlapping segments (s4-s7) mean mutual friends in OSNs. For example, s7 is the intersection of u1, u2, and u3. The overlapping segments might lead to privacy conflicts if some co-owners reject the privacy policy created by the owner.

The proposed CPP is compared with two existing algorithms: naive solution and Hu solution [13]. This comparison uses two metrics based on the research of Hu et al [13]: privacy risk and sharing loss. The privacy risk is the possible degree of harm

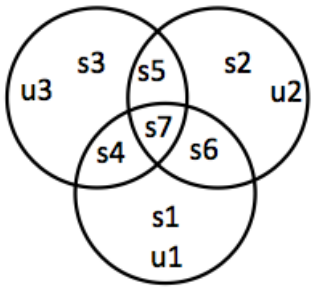


Figure 6.16: Venn Diagram for privacy conflict identification

when the collaborative information is not controlled by the owner and co-owners. The privacy risk can be considered by how sensitive the information is and how wide the spread. The sharing loss is the degree of information not allowed to be spread around. In other words, if the sharing loss is equal to 0, this means no information leak.

The naive solution is a straightforward algorithm for privacy protection. It allows only mutual friends to see collaborative information as shown in Figure 6.16, s7. The Hu solution indicates a trade-off between information sharing and privacy protection by allowing the owner and co-owner to create their own privacy policy. This privacy policy relies on the trust level indicating a set of user names, friendship, or a set of group names and sensitivity level for collaborative information. The privacy conflicts are then identified based on the entire privacy policy. The proposed CPP applies the majority vote concept, where only the owner creates the privacy policy for collaborative information. This policy is voted on by the co-owners. If the vote results show more than half, this collaborative information can be posted. When the co-owner wants to post this information to their friends, family, or other users, the co-owner can change their position to that of owner and create the privacy policy. This experiment is set in the same environment as Hu et al. [13].

- **Results and discussion for measuring performance of the proposed CPP**

  Figure 6.17 depicts different results using three techniques. Each result is measured by using privacy risk and sharing loss.

– **Naive solution**

The naive solution is a simple technique with fast implementation [13]. It offers the highest privacy protection in comparison to the rest, as presented in Figure 6.17(a). This technique grants the mutual friends of u1, u2, and u3 the ability to see the collaborative information and therefore its privacy risk is equal to 0. This can indicate that the owner and co-owners are protected by a strong privacy setting. For sharing loss, the naive solution has the maximum value because there is no way that the collaborative information will be spread outside the set of mutual friends. However, its privacy protection is too strict for practical use. This is not relevant to the objective of OSNs, as they try to create communication and interaction among people through the sharing of information. The owner might want to spread the collaborative information to other readers in addition to the group of mutual friends. From this point of view, the naive solution is considered to offer the highest degree of privacy protection and sharing loss. Another limitation of the naive solution is that it does not consider an instance where the co-owner does not want to share information with any readers, thus using the naive solution might cause this co-owner to lose privacy.

– **Hu solution**

Results from the Hu solution are shown in Figure 6.17(b) and reported in [13]. This sets that u1 and u3 allow the overlapping segments between u1 and u2, and between u2 and u3 to see the collaborative information. It then calculates the resolving score by a combination of the privacy risk and sharing loss in order to compromise between information sharing and privacy protection. Based on the resolving score, a list of readers who can see the collaborative information is designed for all owners and co-owners. Therefore, the privacy risk and sharing loss are balanced by the formulas. The worst case scenario for the Hu solution is the same as for the naive solution, in that only the mutual friends of u1, u2, and u3 can see the information. Hu solution's advantages are that it gives the owner and co-owners an opportunity to participate in the collaborative information. They can create the privacy policy because each

person has different privacy concerns.

In terms of the sharing loss, this solution has more or less the same problem as the naive solution, but the degree is lower. Therefore, the owner cannot share the collaborative information as desired. In terms of privacy risk, even though only overlapping segments of u1 and u2, and u2 and u3 can see the collaborative information, some of them do not know each other. For example, some mutual friends of u1 and u2 are strangers to u3, but they can this collaborative information. However, the degree of privacy risk is lower than in the proposed CPP.

This solution forces everyone to set the privacy policy. This does not accord with the behavior of the creator or owner based on the analysis results in Section 6.4.1 and other research works [91][127]. The creator or owner has difficulty with settings. In addition, it is possible in practical terms that some co-owners might not satisfy current situations with privacy policy. Those co-owners contact the owner to modify the privacy policy, and as a result, it can become an endless problem causing the collaborative information to be out of date when it is posted

– **Proposed CPP**

The proposed CPP uses the majority vote concept to balance between information sharing and privacy protection. The proposed CPP's results can be divided into two types, as depicted in Figure 6.17 (c) and (d). However, the overlapping segments might (not) conflict depending on the vote by the co-owners. For example, u2 and u1 are assumed to be the owner and co-owner respectively. If u2 creates the privacy policy and u1 rejects this policy, s6 and s7 are considered to conflict with u1 and u2. In this case, the proposed CPP assumes that u2 is the owner, and u1 and u3 are the co-owners. In the first type, u1 and u3 accept the privacy policy created by u2, so the value of privacy risk and sharing loss is 0. This indicates that all members including mutual friends of u1 and u3 in the contact list of u2 can see the collaborative information without privacy conflict occurring. In the second type, u1 accepts the

privacy policy created by u2, but u3 rejected it so u2 can spread the collaborative information to all members in the contact list, except those members who are mutual friends of u3. Therefore, the privacy risk is still equal to 0, but the sharing loss is more than 0.

In terms of sharing loss, the owner can spread the collaborative information based on his/her desire. This is consistent with the objective of OSNs to get as many people as possible to communicate and interact with each other through information sharing. This differs from the naive and Hu solutions. In terms of privacy risk, the privacy of the co-owners is still protected even though other users, being members of the owner, can see the collaborative information. Based on the analysis results in Section 6.4.1 and previous research works [83][84], the co-owner will not be too worried if sensitive information leaks to other users, who the co-owners do not have a relationship with because they might not meet each other in the real world.

The proposed CPP realizes the user's behavior when using the privacy setting provided by OSNs, which is hard to understand and boring. As a result, the workload for creating the privacy policy relies on the owner, while the co-owners just consider this policy and make a decision whether or not collaborative information should be posted on OSNs. This also shows that the owner tries to take responsibility for the co-owner's privacy. However, the limitation of the proposed CPP is that the collaborative information might not be up to date due to waiting for the voting results from co-owners. Another limitation is that this research cannot completely solve the problems with robbery, kidnapping, and so on because other users, with whom the co-owners have no relationship, can see the collaborative information although the co-owner rejects the privacy policy created by the owner.

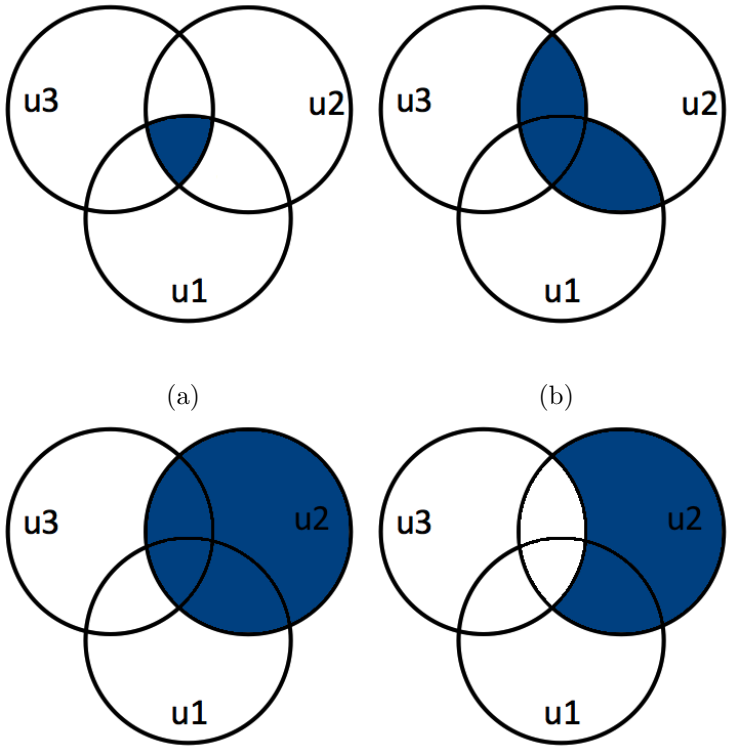


Figure 6.17: (a) Naive solution, (b) Hu solution, (c) the proposed CPP type 1, (d) the proposed CPP type 2

6.5 Dicussion

In this research, the owner attempts to take responsibility for the co-owner's privacy, and therefore the co-owners become part of the decision-making process by using majority vote. Nonetheless, this is still a privacy problem because there might be a minority group of the co-owners who do not want collaborative information posted on OSNs. The co-owners normally have privacy or portrait rights depending on the type of collaborative information, but if the majority vote is applied in the decision making process, these co-owners should respect the group decision. Otherwise, these co-owners should contact and express the negative effect of posting this collaborative information to the owner. However, this research tries to protect privacy by not allowing the mutual friends of the owner and co-owners who accept and reject to see the collaborative information.

There are different sensitivity levels on information arising from cultural differences. In some cultures, text wording and actions in a photo or video affect feelings, image, and traditional cultures. Therefore, seeing the same information can lead to different privacy concerns. In other words, each co-owner makes decisions for their own reasons whether or not he/she allows the collaborative information to be posted on OSNs. For example, an owner in the West views a photo showing kissing in public, but the co-owners in Asia think it is shameful and do not want it spread on OSNs. Another example is that the Miss World organizers confirmed that contestants would not wear bikinis in respect to traditional Muslim cultures during the contest in Indonesia [129]. This indicates that the bikini is a sensitive issue in Muslim countries.

The negotiation of privacy based on culture is also an issue. In the proposed CPP, the owner communicates with co-owners by sending an invitation in order to ask for their permission. Nonetheless, co-owners might not express a real intention to the owner whether or not they want this collaborative information to be posted on OSNs. Many factors influence the co-owner's decision. For instance, if the owner is the co-owner's boss, the co-owner might not wish to cause conflict with the boss by consenting to the collaborative information being posted on OSNs. As another instance, if the co-owners are from collectivist countries like Thailand, China, or Japan and individualist countries like

America, Germany, or Australia, those from the collectivist countries will support each other's answers. Therefore, the owner might not receive the real answer. Furthermore, this negotiation is not done via face-to-face communication, thus when co-owners are in different places to the owner, the emotions of both the owner and co-owners cannot be transmitted via facial expression, body posture, or voice tone. Thus, the owner cannot guess the real meaning from any cues expressed by the co-owners. For this reason, it is necessary to understand the real meaning of the owner and co-owners with different cultures.

6.6 Conclusion

This chapter provides user terms and information definitions, and describes the privacy problem used in this research. The proposed CPP balances collaborative information sharing and privacy protection for the owner and co-owners by using majority vote. It enables the owner to create the privacy policy and the co-owners to make a decision on it. It additionally identifies and provides a suitable solution for those conflicts since at least one co-owner intends to keep the information private. Advantages of the proposed CPP are that the owner and co-owners share collaborative information on OSNs with less privacy concern. The proposed CPP brings about privacy negotiation based on the cultural differences between the owner and co-owners by asking permission. It reduces problems with robbery, defamation, and portrait rights in society.

Chapter 7

Thesis contribution

7.1 Social impact

Nowadays, Online Social Networks (OSNs) are widely used in several domains, such as news reports, business, finding lost family members, and so on. However, there are certain limitations. For example, in the e-commerce domain, companies cannot create a good marketing plan if they only have basic information, such as age group, gender, and so on. Thus, this section aims to provide example approaches to utilize results from this research.

1. **Application for other cultures**

   - **Problem**

     OSNs are nowadays used by many countries in the world, such as America, France, Japan, Thailand, and so on. In the future, OSNs are predicted to continuously grow with more diverse populations. This shows that OSN users are culturally diverse. However, current readers in OSNs are treated equally with the Information Feeding Mechanism (IFM), receiving the same information even though readers in each country have different cultures. For example, Twitter feeds and sorts information by using only timeline. This might satisfy some cultures, but not every culture will like timeline in Twitter. Thus, these readers might not totally consume all information from the IFM. The readers of each culture should be treated individually.

- **Supportive reason**

  The proposed IFM in this research tries to serve information to the reader based on culture, and hence cultural differences are discussed regarding information consumption. Although only Japanese and Thai cultures are investigated, the analysis results can be applied to many countries where the cultural characteristics are quite similar. The cultural characteristics of each country will be similar or different depending on the focus of the dimension. If we consider communication style, the results obtained from Japanese respondents can be applied to Arab countries, Greece, and so on, because these countries are in the same group of high-context cultures as denoted in Figure 7.1. Hofstede's research provides the cultural differences in terms of score and mapping. Therefore, based on Hofstede's uncertainty avoidance and power distance dimensions [41] in Figure 7.2, Thai results can be adopted for Taiwan, Canada, etc. whereas Japanese results can be applied to Poland, South Korea, and so on. Hofstede's research has assisted with this study by saving time in data collection from each country. In addition, if sufficient data is available for each country, it can be applied to the proposed IFM in order to reduce information overload by taking culture into consideration. This is because this data will be analyzed to find suitable features and factors to build the IFM step by step based on culture. Hence, the efficiency of the proposed IFM in reducing information overload relies on cultural differences.

2. **Marketing plan**

- **Problem**

  The use of OSNs in business is currently rising year by year. Brand pages are usually created to promote many products, such as clothing, baby items, children's toys, appliances, and so on. Brand owners usually update or advertise the product information based on their desire without considering customer availability or time suitability. Therefore, it is a fact that information can be spread to all potential customers. However, this information can be annoying to customers because it appears on their Social Network Page (SNP) at

Figure 7.1: High-context cultures and Low-context cultures [9]

inappropriate time. Finally, it is not recognized and ignored by many readers or customers. This indicates that this type of marketing is not efficient in terms of recognition although the brand owner can distribute the advertisement widely. In other words, it is not beneficial for brand owners or customers.

For instance, a brand owner has just obtained a new product. He/she advertises this product via the brand page on Facebook early in the morning or during work time. As that time is for sleeping and working, customers might not see it. Later, this advertisement will be blocked by other information, which has just been fed to the customer's SNP.

- **Supportive reason**

  The proposed IFM of this research helps companies achieve their goal without cost in promoting their products to thousands of people. It controls the information, including posting of the advertisement at the most suitable time, based on the customer's current situation and nationality. Thus, the customer consumes the advertisements with less annoyance and the proposed IFM might help the customer to recognize the product more easily.

Figure 7.2: Hofstede's uncertainty avoidance and power distance dimensions

3. **Reduction of the crime problems**

- **Problem**

  The aim of OSNs is for many users around the world to communicate and interact with each other through information consumption and sharing, and users can take advantage of these services. However, when certain collaborative information is posted on OSNs without asking permission from the co-owners, there is a possibility that this collaborative information causes problems for them in the future. This is because the co-owners know that they are tagged or mentioned on the collaborative information when this information is posted on OSNs. Furthermore, this information can be monitored by any readers and spread rapidly. Thus, it is hard to control. Eventually, the owner and co-owners might unintentionally face crime problems, such as defamation, robbery, and

violation of portrait rights.

- **Robbery**

  Criminals take time to pay attention in observing the victim's activities in each day from the collaborative information which the victim shares, tags, or mentions. They get the exact location of the victims if the information is provided by someone with whom the victim has a relationship or by the victim themselves. They can link the victim's story in order to steal valuables. However, the criminal might not be a stranger but might be someone with whom the victim has a relationship.

  For instance, the victim and her family take a nice trip to Japan. She is tagged on a photo as illustrated in Chapter 6, Figure 6.3. The criminal, who might have a relationship with the victim's friend, knows that she is not at home but on holiday. Thus, the criminal goes to the victim's home and steals valuables.

- **Defamation**

  This problem can occur in OSNs due to the lack of an adequate mechanism to prevent loss of privacy. Certain co-owners may be tagged or mentioned in the collaborative information. This could damage their reputation. In a case where the information is shameful, if it is not true, tagged or mentioned co-owners might have their reputation damaged, leading to misunderstanding or rumor. Even if it is true, certain co-owners might not want this information to be published on OSNs.

- **Portrait right**

  The latest technology means that many people can use a mobile phone or a camera to take photos and record videos everywhere they go. These photos and videos reflect real activities and events. Thus, many creators presently upload photos and the videos on OSNs. It is a fact that these photos and videos might contain not only images of the owner, but co-owners' faces can also appear on them. Therefore, posting such items on OSNs without asking permission from the co-owners might lead to a violation of portrait

rights. For example, a co-owner may be drunk and a photo is taken during a party. When this photo is posted, and the co-owner's family or boss sees it, he/she might have a conflict with this group of people. Therefore, this co-owner has portrait rights to not allow this photo to be posted on OSNs.

- **Supportive reason**

  The proposed Collective Privacy Protection (CPP) attempts to protect the privacy of the owner and co-owners. In circumstances where the owner tries to take responsibility for the co-owners' privacy in collaborative information, the co-owners participate in the decision-making process by using majority vote. The co-owners are notified that their information is being managed by others and decide whether or not such collaborative information should be posted on OSNs. This indicates that allowing co-owners to make a decision can alleviate the problems described above although it cannot fully solve crime problems.

7.2 Academic impact

So far, this research has described details of the proposed IFM and CPP. Not only can this be applied to society, but other research fields can also take advantage of the analysis results of this research. How this research can be used in three main research fields is explained below.

1. **User interface design based on culture**

   - **Problem**

     User interface design requires a good understanding of user needs. In the rush hour, smart information organization will help the reader consume interesting information quickly and completely. Therefore, as well as choosing and placing the elements in the most suitable position on the SNP, this research field should not ignore information organization based on the cultural differences in OSNs. There are three problems with the current user interface of OSNs.

     Firstly, when the retrieved interesting information is ordered at random, some information might not be instantly recognized by the reader [130], since it is

placed in the wrong position. Moreover, ordering in this way may not be appropriate for the small screen of a mobile phone because the reader has to scroll down to see more information. Secondly, current information organization does not allow the reader to control information by selecting a category based on their preferences. So, the reader cannot select specific interesting information to consume on the SNP [130]. Finally, the reader is forced to consume information by reverse chronological order or top stories, so this information organization on the SNP is not consistent with the reader's expectations [130]. Some readers such as Thais expect to see information ordered by their preferences.

- **Supportive reason**

  The proposed IFM allows the reader to specify information based on their preferences at that time. Therefore, it helps the reader to access the required information immediately and is suitable for the small screen size of a mobile phone. Also, the proposed IFM studies cultural preferences in information organization. When the interesting information is retrieved, it is ordered according to the reader's culture. For example, some Japanese readers prefer reverse chronological order to display the information because they can get the latest information. The proposed IFM helps the reader to increase their ability to access and consume information.

2. **Collaborative work based on culture**

   - **Problem**

     Collaboration is working with others to carry out a task and achieve shared goals. However, the goals of a group might not be achieved when each co-worker comes from a different culture and do not understand the cultural characteristics of their co-workers.

   - **Supportive reason**

     The analysis results can increase understanding of the cultural differences among co-workers. Firstly, time in Japanese culture is important. The co-workers, working with Japanese should not arrive late to meetings and should

finish tasks on time. On the other hand, Thai culture views time flexibility as being acceptable in certain situations. Secondly, the word "はい" in Japanese culture means not only "yes" and "ok" but also signals agreement, so co-workers need to consider the context during conversations or make sure they understand the real meaning. Thirdly, for writing styles, the emoticons used by Japanese people to communicate, refer to eye contact, such as (^ v ^), ((^_^)/), and so on. However, emoticons used by Western people use mouth movements to indicate emotional states, such as :-), :-(and so on. Therefore, readers should be careful not to misunderstand certain situations when interpreting the meaning of emoticons. Finally, Japanese and Thais tend to use both non-verbal and verbal communication, and hence co-workers with a low context culture are required to observe facial expressions to support their interpretation.

3. **Negotiation in privacy and business based on culture**

   - **Problem**

     When the owner and co-owners come from different cultures and have not negotiated via face-to-face communication, the owner might misunderstand the co-owner's answer (accept or reject the privacy policy). This is because the owner cannot directly guess the real meaning from any cues, such as facial expression, body language, and so on. In the case of business, nowadays co-operation among countries has increased and business communication is often conducted by e-mail. Therefore, the negotiator needs to understand cultural differences in the country they are dealing with.

   - **Supportive reason**

     In the proposed CPP, even though co-owners from different cultures can control the collaborative information by vote, the analysis results show that many factors influence communication, such as power distance, individualism vs. collectivism, and so on. In the case of power distance [41] which represents inequality in society; subordinates, students, or younger people have to respect their boss, teacher, or elders Therefore, if the co-owner has a lower position

in society, the co-owner's answer might be reliant on a person in a higher position. In another case, if the co-owners are from collectivist countries like Thailand, China, Japan and individualist countries like America, Germany, and Australia, the co-owners from the collectivist countries are likely to support the answer from someone in a similar position to themselves. Therefore, the owner might not receive a true answer from such co-owners.

Chapter 8

Conclusion and Future works

8.1 Conclusion

Technology in the past was poor, and hence conveying information from one to another one might be ineffective. Later, technology was developed and has improved continuously, such as with Internet, telephone, and so on. People can easily communicate with each other across long distances although they are in different places throughout the world. This brings people closer. Nowadays, Online Social Networks (OSNs), i.e. Facebook, Google+, Twitter, etc., have created a major impact on communication and social interaction. OSNs allow users of different races, age groups, and gender around the world, to share any information in the space provided. At the same time, these users can consume large amounts of information anywhere at any time. OSNs are beneficial in several aspects, such as business, charity donations, politics, and so on.

Although OSNs have many advantages, they can lead to an increase in certain problems in relation to cultural differences since the users in OSNs come from many countries around the world as indicated in Figure 1.3. When these users consume and share information on OSNs, the variety and amount of information, or the sensitivity level of each piece of information increases dramatically due to cultural differences. This research focuses on the information overload and loss of privacy problems by cultural considerations in the use of OSNs.

The overload problem in information consumption, arises from cultural patterns of text (i.e. writing style, language, etc.), the large number of users in OSNs, and the large amount of information on OSNs. Therefore, the readers miss interesting or important information and feel confused, anxious, and annoyed when they consume excessive information. Consequently, the reader cannot satisfactorily consume high-quality information. Thus, in this research, the cultural differences of information consumption are studied by using a set of influential features and factors. The readers in OSNs come from different cultures, and thus they might have different criteria when consuming information. Thereafter, a new type of Information Feed Mechanism (IFM) is proposed to reduce the information overload problem by considering cultural differences. The proposed IFM differs from the existing IFM used in OSNs because it uses the same algorithm for serving information to every reader equally. There is no universal IFM for all cultures.

The proposed IFM is more reliable for OSNs readers for practical use because the Naïve Bayes (NB) algorithm offers the highest performance for classification accuracy and the fastest time complexity when compared to the Decision Tree algorithm and the K-Nearest Neighbor algorithm. It can efficiently work with a very large data set, which contains billions of instances in OSNs by using a parallelism concept. Moreover, the set of influential features and factors used in the proposed IFM relates to the cultural differences in OSNs. Accordingly, the proposed IFM helps the readers to reduce the time spent in finding interesting information by providing that which is most suitable based on the reader's current situation and nationality. The retrieved interesting information is ordered by the reader's preference. The proposed IFM is beneficial, not only for readers, but also for marketers in OSNs, since an advertisement can appear on the Social Network Page (SNP) at the most suitable time based on the customer's current situation and nationality. Thus, the reader consumes the advertisements with less annoyance. Even though this study concentrates on the cultural differences of Japanese and Thais, the analysis results can also be applied to other cultures, societies, and businesses.

For the loss of privacy problem in information sharing, privacy concerns are considered to be a crucial problem with OSNs. The current OSNs do not provide a mechanism for

collective privacy management. Only the owner can control the collaborative information. Furthermore, the owner can tag and mention the co-owners in collaborative information without asking their permission. Therefore, the co-owners might not realize their information is being managed by others until the collaborative information is posted on OSNs. It is possible that the collaborative information might leak to unwanted target readers and cause the owner and co-owners loss of privacy.

To address the loss of privacy problem, Collective Privacy Protection (CPP) is proposed for balancing the need for information sharing and privacy protection for the owner and co-owners. The proposed CPP applies a concept of majority vote. It enables the owner to create the privacy policy and the co-owners to make a decision on the privacy policy by vote. The proposed CPP identifies privacy conflicts between the owner and co-owners and provides a suitable solution for those conflicts. It still protects the co-owners' privacy even though only one co-owner rejects the privacy policy. This research analyzes the factor and combination of factors which help to prevent the leakage of information to unwanted target readers and investigates the opinions of co-ownership. Additionally, the performance of the proposed CPP is compared to other research works. By using the proposed CPP, the owner and the co-owners share the collaborative information on OSNs with less privacy concerns because it does not leak to unwanted target readers. The proposed CPP encourages the owner to take responsibility for the co-owners' privacy by asking permission as to whether or not this collaborative information should be posted on OSNs. This means that the owner and co-owners have participated in the control of the collaborative information. The proposed CPP indicates that many factors affect the co-owner's decision during the negotiations, such as power distance, individualism vs. collectivism, and so on. Furthermore, the proposed CPP is beneficial for society by alleviating crime problems such as robbery, defamation, and violation of portrait rights.

Based on this research, the large amount of information, its variety, and the sensitivity levels arising from users with different cultures, can be controlled, thus reducing tension for the users. They can take great advantage from more relaxed information consumption and sharing. Moreover, this can be applied to social and academic aspects, such as

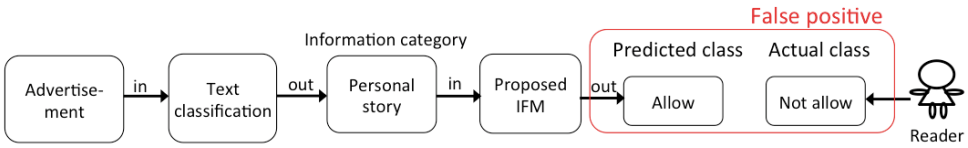
marketing plans, collaborative works based on culture, and so on.

8.2 Future works

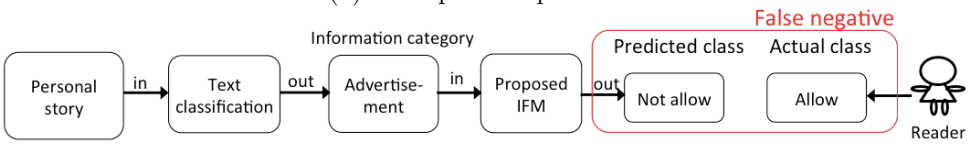
- **Improving performance of the proposed IFM**

  The performance of the proposed IFM could be improved by the further development of text classification. Text classification is used to predict the category of information and is one of the features used in the proposed IFM. Output from text classification can cause *false positive and false negative predictions* in the proposed IFM. For the false positive prediction in Figure 8.1 (a), the actual information is an advertisement, but text classification predicts it as a personal story. The proposed IFM then misunderstands this information to be a personal story, and therefore the proposed IFM uses it together with other features and factors. Finally, the proposed IFM predicts that this information should be shown on the SNP, but the reader does not want to see it. Hence, this information annoys the reader. However, if only a few advertisements appear on the SNP, they do not pose a serious problem. For the false negative prediction in Figure 8.1 (b), the actual information is a personal story and the reader wants to read it but the proposed IFM misunderstands that it is an advertisement, and filters this information, so this information is missed.

  Usually, text classification produces one output, which obtains the highest probability and the proposed IFM uses it. It is quite a risk if the category of information is ambiguous. However, the future plan is to allow the text classification to produce two candidate outputs. If the probabilities of the first and second ranked outputs are close, the proposed IFM will not use the first ranked output. This is because it is possible that the second information is filtered although it is interesting. The proposed IFM in the future will consider the interest level because the current proposed IFM considers that information has two interesting levels (100% for interesting information and 0% for uninteresting information), and these are not enough. Thus, some information which has a high interest level is filtered by the current proposed IFM. Using the interest level helps the reader to avoid missing interesting infor-



(a) False positive prediction



(b) False negative prediction

Figure 8.1: False prediction in the proposed IFM caused by the text classification

mation although the amount of information increases on the SNP. In addition, the reader controls the information by using the interest level.

The future plan involves compacting information of a similar nature to form a single content because many creators in OSNs have the freedom to post any information, and sometimes it is unintentionally similar. Therefore, the reader often consumes the same kind of information and feels bored. Moreover, the reader's SNP contains a large amount of the same information. Compacting similar information helps the reader to save time in consuming the information and avoids the boredom of seeing many similarities on the SNP. Furthermore, it reduces the amount of information on the SNP.

- **Providing privacy protection to support negotiations based on culture**

  Culture is defined by different levels [41][131], ranging from national, professional, organizational, group, and individual levels. Each level indicates different behavior. Although the proposed CPP is good for negotiation based on culture by using individual level analysis, problems with misunderstandings may occur. For example, Figure 8.2 shows people with different nationality levels have different attitudes toward wearing a bikini. Muslim people realize that it is not proper to wear the bikini whereas American people think it is normal. This indicates that creators who have freedom of expression might sometimes be insensitive to others, such as in areas concerning political opinions, religious beliefs, or royal institutions.

Therefore, as shown in Figure 8.3, the plan of this study is to improve privacy protection for negotiation based on culture by using knowledge levels arising from studying privacy preference at national, professional, organizational, and group levels, together with individual behavior. Supporting the national level based on cultural differences is a challenging task because it helps reduce the problems of misunderstanding, not only at the national level, but also for other levels in the negotiations based on culture. Thereafter, the knowledge level is to be used to conduct sensitivity levels of information for providing privacy protection together with the proposed CPP. In the case of private information, the creator is reminded as to whether or not the information being posted could be sensitive to other readers. With collaborative information, the sensitivity level of the information helps to improve the negotiations based on culture when it is sensitive to the co-owner. However, the co-owner still allows this information to be posted on OSNs because he/she does not want to conflict with other co-owners.

Figure 8.2: Different attitudes towards wearing a bikini at a national level

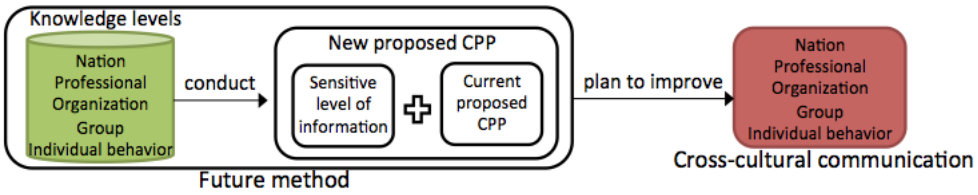


Figure 8.3: Privacy protection to support the negotiations based on culture in future research

Bibliography

[1] The Evolution of Mass Communication.
http://thecontentguy.wordpress.com/2011/03/23/the-evolution-of-mass-communication/.

[2] Facebook passes 1.19 billion monthly active users 874 million mobile users and 728 million daily users.
http://thenextweb.com/facebook/2013/10/30/facebookpasses119billionmonthly-activeusers874millionmobileusers728milliondailyusers/#rLXRr.

[3] GooglePlus wikipedia.
http://en.wikipedia.org/wiki/File:Google

[4] LinkedIn, The modified design of the profile page.
http://japan.cnet.com/news/service/35023149/.

[5] C. Yang, J. Sun, and Z. Zhao, "Personalized recommendation based on collaborative filtering in social network," in *Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on*, vol. 1, pp. 670–673, 2010.

[6] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. in Artif. Intell.*, vol. 2009, pp. 4:2–4:2, 2009.

[7] Victim of Secret Dorm Sex Tape Posts Facebook Goodbye, Jumps to His Death.
http://abcnews.go.com/US/victim-secret-dorm-sex-tape-commits-suicide/story?id=11758716.

[8] Teachers' Virtual Lives Conflict With Classroom.
http://abcnews.go.com/TheLaw/story?id=4791295.

[9] E. Wurtz, "Intercultural communication learning a model of a web use on web sites a cross-cultural analysis of web sites from high-context cultures and low-context cultures," *Journal of Computer-Mediated Communication*, vol. 11, no. 1, pp. 274–299, 2005.

[10] Facebook.
http://www.facebook.com.

[11] Google+.
https://plus.google.com.

[12] Twitter.
https://twitter.com.

[13] H. Hu, G.-J. Ahn, and J. Jorgensen, "Detecting and resolving privacy conflicts for collaborative data sharing in online social networks," in *Proceedings of the 27th Annual Computer Security Applications Conference*, ACSAC '11, (New York, NY, USA), pp. 103–112, ACM, 2011.

[14] Making a Social Impact: Innovating with Facebook to Engage and Inspire.
https://developers.facebook.com/blog/post/2012/10/04/makingasocialimpact-innovatingwithfacebooktoengageandinspire/.

[15] List of virtual communities with more than 100 million active users.
http://en.wikipedia.org/wikiList_of_virtual_communities_with_more_than_100_million_active_users.

[16] Facebook Reports Third Quarter 2013 Results.
http://investor.fb.com/releasedetail.cfm?ReleaseID=802760.

[17] A. G. Schick, L. A. Gordon, and S. Haka, "Information overload: A temporal approach," *Accounting, Organizations and Society*, vol. 15, no. 3, pp. 199 – 220, 1990.

[18] K. Bontcheva, G. Gorrell, and B. Wessels, "Social media and information overload: Survey results," *CoRR*, vol. abs/1306.0813, 2013.

[19] M. J. Eppler and J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines," *The Information Society*, vol. 20, no. 5, pp. 325–344, 2004.

[20] J. B. Strother, J. M. Ulijn, and Z. Fazal, *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. Jonh Wiley & Sons, Inc., 2012.

[21] With 1.5 billion image uploads per week, Google+ focuses in on photography. http://www.engadget.com/2013/10/29/googleplusphotoscreateanaudience/.

[22] M. B. Brewer, "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis," in *Psychological Bulletin*, vol. 86, pp. 307–324, 1979.

[23] A. C. Squicciarini, M. Shehab, and J. Wede, "Privacy policies for shared content in social network sites," *The VLDB Journal*, vol. 19, pp. 777–796, Dec. 2010.

[24] H. K. Tsoi and L. Chen, "From privacy concern to uses of social network sites: A cultural comparison via user survey," in *PASSAT/SocialCom 2011*, (Hong Kong), pp. 457–464, 2011.

[25] Facebook EdgeRank and GraphRank Explained. http://blog.involver.com/2011/10/25/facebook-edgerank-and-graphrank-explained.

[26] GooglePlus:View and filter your Home page with circles. https://support.google.com/plus/answer/1269165?hl=en/.

[27] P. Lops, M. de Gemmis, G. Semeraro, F. Narducci, and C. Musto, "Leveraging the linkedin social network data for extracting content-based user profiles," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, (New York, NY, USA), pp. 293–296, ACM, 2011.

[28] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-based filtering in on-line social networks," in *The international ECML/PKDD conference on Privacy and security issues in data mining and machine learning*, PSDML'10, (Berlin, Heidelberg), pp. 127–140, Springer-Verlag, 2011.

[29] T. Paek, M. Gamon, S. Counts, D. M. Chickering, and A. Dhesi, "Predicting the importance of newsfeed posts and social network friends.," in *AAAI* (M. Fox and D. Poole, eds.), AAAI Press, 2010.

[30] T. N. Dinh, Y. Shen, and M. T. Thai, "The walls have ears: optimize sharing for visibility and privacy in online social networks," in *CIKM*, pp. 1452–1461, 2012.

[31] B. Carminati, E. Ferrari, and A. Perego, "Rule-based access control for social networks," in *Proceedings of the 2006 International Conference on On the Move to Meaningful Internet Systems: AWeSOMe, CAMS, COMINF, IS, KSinBIT, MIOS-CIAO, MONET - Volume Part II*, OTM'06, (Berlin, Heidelberg), pp. 1734–1744, Springer-Verlag, 2006.

[32] P. A. Rutledge, *The truth about profiting from social networking.* Person education, Inc., 2008.

[33] I. B. Dhia, "Access control in social networks: A reachability-based approach," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, (New York, NY, USA), pp. 227–232, ACM, 2012.

[34] K. Bontcheva, G. Gorrell, and B. Wessels, "Social media and information overload: Survey results," *CoRR*, vol. abs/1306.0813, 2013.

[35] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi, "Talking in circles: Selective sharing in google+," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 1065–1074, ACM, 2012.

[36] L. Haddon, *Information and communication technologies in everyday life.* Berg Publishers, 2004.

[37] Social Media Active Users 2013.
https://dustn.tv/activeusers2013/.

[38] GooglePlus:View and filter your Home page with circles.
https://support.google.com/plus/answer/1269165?hl=en/.

[39] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web*, WWW '10, (New York, NY, USA), pp. 591–600, ACM, 2010.

[40] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *4th ACM conference on recommender systems*, RecSys '10, (New York, USA), pp. 199–206, 2010.

[41] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival*. New York: McGraw-Hill, 2010.

[42] M. Minkov and G. Hosftede, *Cross-cultural analysis: the science and art of comparing the world's modern societies and their cultures*. Thousand Oaks CA: Sage Publications, 2013.

[43] C. Kluckhon, *The Study of Culture*. in: Lerner / Lasswell (eds.): The policy science, Stanford: Stanford University Press, 1951.

[44] A. L. Kroeber and T. Parsons, "The concept of culture and of social system," *American Sociological Review*, 1958.

[45] L. White, *The Evolution of Culture*. N.Y.: McGraw-Hill, 1959.

[46] G. Hofstede, *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks CA: Sage Publications, 2001.

[47] National cultural dimensions.
http://geerthofstede.com/nationalculture.html.

[48] R. Inglehart and W. E. Baker, "Modernization, cultural change, and the persistence of traditional values," *American Sociological Review*, 2000.

[49] T. Lenartowics and K. Roth, "Does subculture within a country matter? a cross-culture study of motivational domains and business performance in brazil," *Journal of International Business Studies*, 2001.

[50] N. A. Boyacigiller, J. Kleinberg, M. E. Phillips, and S. A. Sackmann, "Conceptualizing culture : elucidating the streams of research in international cross-cultural management," *Theoretical foundations, critiques & developments.*, 2009. Aus: Handbook for international management research.

[51] R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, and V. Gupta, *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks CA: Sage Publications, 2004.

[52] Confucius and Confucianis in Japanese Art and Culture. http://www.onmarkproductions.com/html/japanese-confucianism.html.

[53] J. Lu, K. L. Chin, J. Yao, J. Xu, and J. Xiao, "Cross-cultural education: Learning methodology and behaviour analysis for asian students in it field of australian universities," in *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103*, ACE '10, (Darlinghurst, Australia, Australia), pp. 117–126, Australian Computer Society, Inc., 2010.

[54] J. Lu, K. Chin, J. Yao, J. Xu, and J. Xiao, *Survey Analysis on Cross Culture Learning*. UTS Teaching and Learning Forum 2008, 2008.

[55] P. Jing and G. Rong, "Exploring cross-cultural factors affecting mobile commerce adoption: Theories and empirical comparison between u.s and china," in *Information Management, Innovation Management and Industrial Engineering (ICIII), 2010 International Conference on*, vol. 3, pp. 324–327, 2010.

[56] C. L. Sia, Y. Shi, C. H. Tan, J. Wei, J. Yang, and N. Wang, "Leveraging social grouping for organizational endorsement in mobile commerce across cultures: Transforming outgroups into ingroups," in *Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR), 2010 Ninth International Conference on*, pp. 487–492, 2010.

[57] M. Quiros-Ramirez and T. Onisawa, "Assessing emotions in a cross-cultural context," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, pp. 2967–2972, 2012.

[58] A. Vasalou, A. N. Joinson, and D. Courvoisier, "Cultural differences, experience with social networks and the nature of true commitment in facebook," *Int. J. Hum.-Comput. Stud.*, vol. 68, pp. 719–728, Oct. 2010.

[59] Y. Kim, D. Sohn, and S. M. Choi, "Cultural difference in motivations for using social network sites: A comparative study of american and korean college students," *Comput. Hum. Behav.*, vol. 27, no. 1, pp. 365–372, 2011.

[60] A. Ratikan and M. Shikida, "A study of cross-culture for a suitable information feeding in online social networks," in *HCI (21)*, pp. 458–467, 2013.

[61] G. Bradley, *Social and community informatics Human on the net.* Routledge, 2006.

[62] BusinessDictionary.com. http://www.businessdictionary.com/definition/information-overload.html.

[63] K. H. Koroleva, Ksenia and O. Gunther, ""stop me!"-exploring information overload on facebook," in *16th Americas Conference on Information Systems*, 2010.

[64] K. Koroleva and A. Bolufe-Rohler, "Reducing information overload: Design and evaluation of filtering & ranking algorithms for social networking sites," in *20th European Conference on Information Systems*, 2012.

[65] M. Grineva and M. Grinev, "Information overload in social media streams and the approaches to solve it," *21st International World Wide Web Conference, Lyon, France*, 2012.

[66] Q. Jones, G. Ravid, and S. Rafaeli, "Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration," 2004.

[67] N. Zeldes, D. Sward, and S. Louchheim, "Infomania: Why we can't afford to ignore it any longer," *First Monday*, vol. 12, no. 8, 2007.

[68] Information filtering system. http://http://en.wikipedia.org/wiki/Information_filtering_system.

[69] R. G. Tingshao Zhu and G. Haubl, *Learning a Model of a Web User''s Interests.* Springer Berlin Heidelberg, 2003.

[70] S. Nakamura and K. Tanaka, "Temporal filtering system to reduce the risk of spoiling a user's enjoyment," in *Proceedings of the 12th International Conference on Intelligent User Interfaces*, IUI '07, (New York, NY, USA), pp. 345–348, ACM, 2007.

[71] S. Loeb and E. Panagos, "Information filtering and personalization: Context, serendipity and group profile effects," in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pp. 393–398, 2011.

[72] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *Internet Computing, IEEE*, vol. 7, no. 1, pp. 76–80, 2003.

[73] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 6, pp. 734–749, 2005.

[74] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 1, pp. 56–69, 2004.

[75] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, (New York, NY, USA), pp. 175–186, ACM, 1994.

[76] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331–370, Nov. 2002.

[77] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *4th ACM conference on recommender systems*, RecSys '10, (New York, USA), pp. 199–206, 2010.

[78] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," 1999.

[79] B. N. Miller, J. A. Konstan, and J. Riedl, "Pocketlens: Toward a personal recommender system," *ACM Trans. Inf. Syst.*, vol. 22, pp. 437–476, July 2004.

[80] E. Bergeron, "The difference between security and privacy," in *Joint workshop on mobile web privacy WAP forum & World Wide Web consortium*, 2000.

[81] D. Rosenblum, "What anyone can know: The privacy risks of social networking sites," *Security Privacy, IEEE*, vol. 5, no. 3, pp. 40–49, 2007.

[82] X. Chen and S. Shi, "A literature review of privacy research on social network sites," in *Multimedia Information Networking and Security, 2009. MINES '09. International Conference on*, vol. 1, pp. 93–97, 2009.

[83] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite, "Computer networks as social networks: Collaborative work, telework, and virtual community," *Annual Review of Sociology*, vol. 22, no. 1, pp. 213–238, 1996.

[84] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, WPES '05, (New York, NY, USA), pp. 71–80, ACM, 2005.

[85] B. Carminati, E. Ferrari, and A. Perego, "Rule-based access control for social networks," in *Proceedings of the 2006 International Conference on On the Move to Meaningful Internet Systems: AWeSOMe, CAMS, COMINF, IS, KSinBIT, MIOS-CIAO, MONET - Volume Part II*, OTM'06, (Berlin, Heidelberg), pp. 1734–1744, Springer-Verlag, 2006.

[86] K. K.Gollu, S. Saroiu, and A. Wolman, "A social networking-based access control scheme for personal content," in *In proceeding of the 21st ACM Symposium on operating systems principles*, SOSP'07- Work-in-Progress Session, 2007.

[87] M. Hart, R. Johnson, and A. Stent, "More content-less control: Access control in the web 2.0," in *In IEEE Web 2.0 privacy and security workshop*, 2007.

[88] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook friends: social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[89] H. Hu, G.-J. Ahn, and J. Jorgensen, "Enabling collaborative data sharing in google+," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, pp. 720–725, 2012.

[90] T. N. Dinh, Y. Shen, and M. T. Thai, "The walls have ears: optimize sharing for visibility and privacy in online social networks," in *CIKM*, pp. 1452–1461, 2012.

[91] Q. Li, J. Li, H. Wang, and A. Ginjala, "Semantics-enhanced privacy recommendation for social networking sites," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pp. 226–233, 2011.

[92] F. Adu-Oppong, C. Gardiner, A. Kapadia, and P. Tsang, "Social circles: Tackling privacy in social networks," in *in Symposium on usable privacy and security (SOSP)*, p. 2, 2008.

[93] E. T. Hall, *Beyond culture*. Anchor Books, 1976.

[94] J. Lu, K. L. Chin, J. Yao, J. Xu, and J. Xiao, "Cross-cultural education: Learning methodology and behaviour analysis for asian students in it field of australian universities," in *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103*, ACE '10, (Darlinghurst, Australia, Australia), pp. 117–126, Australian Computer Society, Inc., 2010.

[95] G. Hofstede, *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thounsand Oaks, London: Sega Publications, 2001.

[96] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and tweet: Experiments on recommending content from information streams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, (New York, NY, USA), pp. 1185–1194, ACM, 2010.

[97] J. Chen, R. Nairn, and E. Chi, "Speak little and well: Recommending conversations in online social streams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 217–226, ACM, 2011.

[98] K. Lerman, "Social networks and social information filtering on digg," *CoRR*, vol. abs/cs/0612046, 2006.

[99] Google calendar.
https://www.google.com/calendar‘‘.

[100] T. OKAZAKI, "Punctuality: Japanese business culture, railway service and coordination problem," *INTERNATIONAL JOURNAL OF ECONOMICS AND FINANCE STUDIES*, 2012.

[101] Muslims and Punctuality.
http://www.onislam.net/english/reading-islam/living-islam/personal-stories/in-their-own-words/452660-islam-and-the-importance-of-punctuality.html.

[102] 2012 has been a big year on the Japanese social-media scene.
https://www.japantimes.co.jp/life/2012/12/19/digital/2012-has-been-a-big-year-on-the-japanese-social-media-scene/#.U1h5ib-gdSU.

[103] Twitter is huge in Japan bigger than Facebook, actually.

[104] Facebook Ads.
https://www.facebook.com/settings?tab=ads/.

[105] I. H. Witten, E. Frank, and M. A. Hall, *Data mining : practical machine learning tools and techniques*. Morgan Kaufmann, 2011.

[106] D. Quinn, L. Chen, and M. Mulvenna, "Does age make a difference in the behaviour of online social network users?," in *Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing*, pp. 266–272, 2011.

[107] U. Pfeil, R. Arjan, and P. Zaphiris, "Age differences in online social networking ‘‘ a study of user profiles and the social capital divide among teenagers and older users in myspace," *Computers in Human Behavior*, vol. 25, no. 3, pp. 643 – 654, 2009.

[108] E. Hargittai, "Whose space? differences among users and non-users of social network sites," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 276–297, 2007.

[109] Selected sections from the country file of Thailand.
http://www.ncl.ac.uk/ecls/assets/documents/pdf/countryfiles/Thailand.pdf.

[110] A. Ratikan and M. Shikida, "Feature selection based on audience's behavior for information filtering in online social networks," in *7th International Conference on Knowledge, Information and Creativity Support Systems*, (Australia), pp. 81–88, 2012.

[111] R. J. Davies and O. Ikeno, *The Japanese mind: understanding contemporary Japanese culture*. Boston ; Tokyo : Tuttle Pub, 2002.

[112] Facebook developer.
http://developers.facebook.com/.

[113] Facebook Query Langauge.
https://developers.facebook.com/docs/technical-guides/fql.

[114] F. Sebastiani and C. N. D. Ricerche, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, pp. 1–47, 2002.

[115] Document classification.
http://en.wikipedia.org/wiki/Document_classification.

[116] I. Androutsopoulos, G. Paliouras, and E. Michelakis, "Learning to filter unsolicited commercial e-mail," tech. rep., Athens University of Economics and Business, Athens, Greece, 2004.

[117] Decision trees and leaf counting.
http://web.engr.illinois.edu/∼jeffe/teaching/algorithms/.

[118] L. Zhou, H. Wang, and W. Wang, "Parallel implementation of classification algorithms based on cloud computing environment," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 10, no. 5, pp. 1087–1092, 2012.

[119] N. Amado, J. Gama, and F. Silva, "Parallel implementation of decision tree learning algorithms," in *Progress in Artificial Intelligence*.

[120] C. Zhang, F. Li, and J. Jestes, "Efficient parallel knn joins for large data in mapreduce," in *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12, (New York, NY, USA), pp. 38–49, ACM, 2012.

[121] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: Interactive topic-based browsing of social status streams," in *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, (New York, NY, USA), pp. 303–312, ACM, 2010.

[122] L. Katz, *Negotiating International Business*. Booksurge, 2007.

[123] R. B. Harris and D. Paradice, "An investigation of the computer-mediated communication of emotions," in *Journal of Applied Sciences Research*, vol. 3, pp. 2081–2090, 2007.

[124] B. Kavanagh, "A cross-cultural analysis of japanese and english non-verbal online communication: the use of emoticons in weblogs," in *Intercultural Communication Studies*, vol. 19, pp. 65–80, 2010.

[125] A. Ratikan and M. Shikida, "Privacy protection based privacy conflict detection and solution in the online social networks," in *16th international conference on human-computer interaction*, (Greece), p. accept, 2014.

[126] L. Banks and S. Wu, "All friends are not created equal: An interaction intensity based approach to privacy in online social networks," in *Computational Science and Engineering, 2009. CSE '09. International Conference on*, vol. 4, pp. 970–974, 2009.

[127] L. Church, J. Anderson, J. Bonneau, and F. Stajano, "Privacy stories: Confidence in privacy behaviors through end user programming," in *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, (New York, NY, USA), pp. 20:1–20:1, ACM, 2009.

[128] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *Security Privacy, IEEE*, vol. 3, no. 1, pp. 26–33, 2005.

[129] Miss World cuts famous bikini contest after Muslim protests in Indonesia. http://nypost.com/2013/06/06/missworldcutsfamousbikinicontestaftermuslim-protestsinindonesia/.

[130] ISO 9241-12:1998(en) Ergonomic requirements for office work with visual display terminals (VDTs)-Part 12: Presentation of information. https://www.iso.org/obp/ui/#iso:std:iso:9241:-12:ed-1:v1:en.

[131] E. Karahanna, J. R. Evaristo, and M. Srite, "Levels of culture and individual behavior: An integrative perspective," vol. 03, pp. 1–20, J. Global Inform. Manag., 2005.

Publications

International journal

[1] Arunee Ratikan, Mikifumi Shikida, "Culture Based Preference for the Information Feeding Mechanism in Online Social Networks," IEICE Transactions on Information and Systems, Vol. E97-D, No.04.

International conferences

[2] Arunee Ratikan, Mikifumi Shikida, "Privacy Protection Based Privacy Conflict Detection and Solution in the Online Social Networks," 2nd International Conference on Human Aspects of Information Security, Privacy, and Trust, LNCS 8533, pp. 433-445, Springer International Publishing, Heraklion, Greece, 2014.

[3] Arunee Ratikan, Mikifumi Shikida, "A Study of Cross-Culture for a Suitable Information Feeding in Online Social Networks," 5th International Conference on Cross-Cultural Design. Cultural Differences in Everyday Life, USA, LNCS 8024, pp.458-467, Springer Berlin Heidelberg, Las Vegas , USA, 2013.

[4] Arunee Ratikan, Mikifumi Shikida, "Feature Selection Based on Audience's Behavior for Information Filtering in Online Social Networks," 7th International Conference on Knowledge, Information and Creativity Support Systems, pp.81-88, Melbourne, Australia, 2012.